

STIC-Biotech/ChemLib

174948

From: Ramirez, Delia
Sent: Thursday, December 22, 2005 4:38 PM
To: STIC-Biotech/ChemLib
Subject: 10/602219

Hi,

I would like to request the following searches:

1. SEQ ID NO: 11 and 12 in the nucleic acid databases (commercial and interference)
2. SEQ ID NO:12 in the protein databases (commercial and interference)

Thank you very much,

Delia M. Ramirez, Ph.D.
Patent Examiner
Recombinant Enzymes-Art Unit 1652
USPTO
400 Dulany Street, Remsen Bldg., 2D74, Mail room 2C70
Alexandria, VA 22314
(571) 272-0938
delia.ramirez@uspto.gov

RECEIVED
DEC 22 2005
1617

Searcher: DiNatale
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: _____
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 26, 2005, 10:52:16 ; Search time 233 Seconds
(without alignments)
3211.817 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MOLRNPEHLGICALRFLA.....RLRSHNPVTGVLLQLENTM 421

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool.p/US10602219/runat.23122005.151142.17165/app.query.fasta_1.583
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MAIRX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602219 @CGN 1.1.290 @runat.23122005.151142.17165 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2.6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2.6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2.6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2.6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2.6/ptodata/1/ina/H COMB.seq.*
6: /cgn2.6/ptodata/1/ina/PCITUS COMB.seq.*
7: /cgn2.6/ptodata/1/ina/pp COMB.seq.*
8: /cgn2.6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2.6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	1266	3	US-09-993-059-11
2	2314	100.0	1266	3	US-10-103-327-11
3	2314	100.0	1278	3	US-09-993-059-7
4	2314	100.0	1278	3	US-10-103-327-7
5	2314	100.0	1284	3	US-09-993-059-13
6	2314	100.0	1284	3	US-10-103-327-13
7	2314	100.0	1290	3	US-09-993-059-3
8	2314	100.0	1290	3	US-10-103-327-3
9	2314	100.0	1296	3	US-09-993-059-9

10	2314	100.0	1296	3	US-10-103-327-9
11	2314	100.0	1308	3	US-09-993-059-5
12	2314	100.0	1308	3	US-10-103-327-5
13	2314	100.0	1343	3	US-08-928-881-18
14	2314	100.0	1343	3	US-09-543-921-18
15	2314	100.0	1343	3	US-09-266-014-3
16	2314	100.0	1343	3	US-09-491-759-18
17	2314	100.0	1393	2	US-07-602-824A-1
18	2314	100.0	1393	2	US-07-983-451-1
19	2314	100.0	1393	2	US-08-261-577-6
20	2293	99.1	1254	3	US-09-993-059-15
21	2293	99.1	1254	3	US-10-103-327-15
22	2293	99.1	1272	3	US-09-993-059-17
23	2293	99.1	1272	3	US-10-103-327-17
24	2293	99.1	8234	3	US-09-626-127-14
25	2293	99.1	8234	3	US-09-993-059-34
26	2293	99.1	8234	3	US-10-103-327-34
27	2234	96.5	1304	9	5179023-3
28	2231	96.4	1233	3	US-09-993-059-21
29	2231	96.4	1233	3	US-10-103-327-21
30	2227	96.2	1215	3	US-09-993-059-19
31	2227	96.2	1215	3	US-10-103-327-19
32	2178	94.1	1214	3	US-09-023-655-1001
33	2152	93.0	1170	3	US-09-176-666-49
34	2152	93.0	1173	3	US-09-176-666-48
35	2152	93.0	1176	3	US-09-176-666-47
36	2152	93.0	1179	3	US-09-176-666-46
37	2152	93.0	1182	3	US-09-176-666-45
38	2152	93.0	1188	3	US-09-176-666-44
39	2152	93.0	1194	3	US-09-176-666-43
40	2152	93.0	1197	3	US-08-928-881-25
41	2152	93.0	1197	3	US-09-543-921-25
42	2152	93.0	1197	3	US-09-266-014-5
43	2152	93.0	1197	3	US-09-491-759-25
44	2147	92.8	1167	3	US-09-176-666-50
45	2142	92.6	1164	3	US-09-176-666-51

ALIGNMENTS

RESULT 1
US-09-993-059-11
; Sequence 11, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUNAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-09-993-059-11

Alignment Scores:	8.08e-252	Length:	1266
Pred. No.:	2314.00	Matches:	421
Score:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-10-602-219-12 (1-421) x US-09-993-059-11 (1-1266)

```
QY      1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db      1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTGGCC 60
QY     21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db      61 CTCGTTTCTGGGACATCCTGGGCTAGAGCACCTGGACAATGGATTGGCAAGGAGCCT 120
QY     41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db     121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180
QY      61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db     181 GATTCTCGCATCAGTGAAGAGCTTCCTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY      81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db     241 TGAAGGATGCAGGTTATGATACCTCTGCATTGATGACTGTGGATGCTCCCCAAAGA 300
QY     101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db     301 GATTCAAGAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCTCATGGGATTGCCAGCTA 360
QY     121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db     361 GCTAAATTATGTTACAGCAAGAGACTGAAGCTAGGAGTTATGCAATGTTGAAATAAA 420
QY     141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db     421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACACTACGACATTGATGCCCCAGACCTTTGCT 480
QY     161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db     481 GACTGGGAGTAGACTGCTGAAAATTGATGGTGTACTGTGACAGATTGGAAAAATTG 540
QY     181 AlaAspGlyTyrIlyshisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db     541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAAAGCATTGTGTAC 600
QY     201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db     601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTCAAAGCCCAATTATACAGAAATCCGA 660
QY     221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db     661 CAGTACTGCAATCACTGGCGAAAATTTTGCTGCAATTGATGATTCCTCGGAAAAGTATAAAG 720
QY     241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db     721 AGTATCTTGACCTGGACATCTTTTACCAGGAGAAATTTGATGTTGCTGACCAGGG 780
QY     261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db     781 GGTGGAAATGACCCAGATATGTTAGTGAITGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
QY     281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db     841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCCTTTATTATGCTAATGACCTC 900
QY     301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db     901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATGAAGGAGCTAATTGGCATCAAT 960
QY     321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db     961 CAGGACCCCTTGGCAAGCAAGGGTACCAGCTTAGACGGGAGACAACCTTTGAAGTGTGG 1020
QY     341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db    1021 GAACGACCTCTCTCAGGCTTAGCGCTGTAGCTATGATGATAAAACCCGCGAGAGATTGGT 1080
```

```
QY     361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db    1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAACTCT 1140
QY     381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db    1141 GCCTGCTTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTCTTATGAATGGACT 1200
QY     401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db    1201 TCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAAACACA 1260
QY     421 Met 421
Db    1261 ATG 1263
RESULT 2
US-10-103-327-11
; Sequence 11, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-103-327-11
Alignment Scores:
Pred. No.:      8,08e-252      Length:      1266
Score:          2314.00      Matches:     421
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:              3          Gaps:          0
US-10-602-219-12 (1-421) x US-10-103-327-11 (1-1266)
QY      1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db      1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTGGCC 60
QY     21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db     61 CTCGTTTCTGGGACATCCTGGGCTAGAGCACCTGGACAATGGATTGGCAAGGAGCCT 120
QY     41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db     121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180
QY     61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db     181 GATTCTCGCATCAGTGAAGAGCTTCCTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY     81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db     241 TGAAGGATGCAGGTTATGATGATACCTCTGCATTGATGACTGTGGATGCTCCCCAAAGA 300
QY     101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
```



```
Db 301 GATTTCAGAGGAGAGCTTCAGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAntyValHisSerLysGlyLeuLysLeuGlyLysLeuValHisSerLys 140
Db 361 GCTAATTAATGTTACAGCAGAGGAGCTAGAGCTAGGAGTTATGCGATTTGGAAATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCTCGGAGTTTGGTACTACGACATTCATGATGCGCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTCTAAATTTGATGGTTTGTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCAGCATGCTCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTAACAGAAATCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCATTTGATGATTCCTCGAAAGTATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGAGCTGACATCTTTTAAACAGGAGAGAAATTTGTGATTTGCTGGACAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTATGGCACTTTGGCCCTCAGCTGGATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaLalaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCCCAAGCTCTCTCTCAGGATAAGGAGTAAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGACAGCAGGCTACAGCTTAGCAGGAGAGCAACTTTGAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGCAGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTATACCATCGACATGCTTCTCTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCTGTGAAAGAGAGCTAGGTTTCTATGATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAGTACATAAATCCACAGSCACTGTTTGTCTTCAGCTAGAAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
```

RESULT 3

```
US-09-993-059-7
; Sequence 7, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
```

```
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-09-993-059-7

Alignment Scores:
Pred. No.: 8,19e-252 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-09-993-059-7 (1-1278)
```

```
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuAtpPheLeuAla 20
Db 1 ATGCAGCTGAGGAGACCCAGAACTACATCTGGGCTGGCGCTTGGCTTTCCTTGCGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCATGTGCACATGTGATGGCAAGGAGCGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGTCAGAGAGGCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCTGCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGATGACAGGTTATGATGACTCTGCAATGATGACTGTTGGATGGCTCCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTAATGTTACAGCAAGAGCTAGAGCTAGGAGTTATGAGATGATTTGGAAATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCTCGGAGTTTGGTACTACGACATTCATGATGCGCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTCTAAATTTGATGGTTTGTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCAGCATGCTCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTAACAGAAATCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
```

Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCAATTTGATGATCTCTGGAAGATATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACGAGG 780
QY 261 GlyTAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGATGATTGGCACTTTGGCTCAGCTGGAAATCAGCNA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTATGTCATAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTTCAGGATAAGGACGTAAATGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGCAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCCTGTAAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGlnTrpThr 400
Db 1141 GCTTGCTTATCACAGCTCTCCTCTGTAAGAGGAGCTAGGCTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAAGTCACATAATCCACAGGCACGTGTTTGTTCAGCTAGAAAACACA 1260
QY 421 Met 421
Db 1261 ATG 1263

RESULT 4

US-10-103-327-7
; Sequence 7, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIORITY FILING DATE: 2002-03-20
; PRIOR FILING DATE: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-10-103-327-7

Alignment Scores:

Pred. No.: 8.19e-252 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-10-602-219-12 (1-421) x US-10-103-327-7 (1-1278)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGACGCTGAGGAACCCAGAACTACATCTGGCTGCGGCTTGGCTTCGCTTCCTCGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGACACTGGCAATGGATTGGCAAGSACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGTGTCGAGGAAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAGGATGCAGGTTATGAGTACCTCTGCATTTGATGACTGTGTGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAAGGCAGACACTTCAGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTACAGCAGAAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTCGCAGGCTTCCCTGGGAGTTTGGTACTACGACATTTGATGTCGCCAGACCTTTGT 480
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGATTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTATAGCACATGCTCTTGGCCCTGATAGGACTGGCAGAACTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCACTTGATGATTCCTGGAAGATATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACGAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTATGTCATAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTTCAGGATAAGGACGTAAATGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluLeuGly 360
Db 1021 GAACGACCTCTCTAGGCTTAGCTGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080

QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGAGAGCTAGGCTTCTATGAATGGACT 1200

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGTTAAGAAGTACATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260

QY 421 Met 421

Db 1261 ATG 1263

RESULT 5

US-09-993-059-13

; Sequence 13, Application US/09993059

; Patent No. 6887696

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KOMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN

; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/09/993,059

; CURRENT FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13

; LENGTH: 1284

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(1284)

US-09-993-059-13

Alignment Scores:

Pred. No.: 8,25e-252 Length: 1284
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-602-219-12 (1-421) x US-09-993-059-13 (1-1284)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGACTACATCTGGGCTGGCGCTTGGCTTCCTTCCTGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGATGGCAAGGAGCGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTTCATGTGCAACTTTCAGCTGGCAGGAAGGCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAGCTTTCATGGAGATGGCAGACTATGGTCTCAGAGGC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspCysTrpMetAlaProGlnArg 100

Db 241 TGGAAGGATGAGGTTATGAGTACCTCTGCAATTGATGACTCTGTTGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120

Db 301 GATTGAGAGGAGGACATTCAGGAGAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaIleTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140

Db 361 GCTAATTATGTTTACAGCAAGAGGCTGAAGCTAGGAGTTTATGCAAGATGTTTGGAAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160

Db 421 ACCTGGCGAGGCTTCCCTGGGAGTTTTGGATCTACGACATTCATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180

Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200

Db 541 GCAGATGGTTATAAGCACATGCTCTGGCCCTGTAATAGGACTGGCAGAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220

Db 601 TCCTGTGAGTGGCCTCTTTATATGCGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240

Db 661 CAGTACTGCATCACTGGCGAAATTTTCTGCACATTCATGATTCCTGGAAAGATATAAG 720
QY 241 SerLeuLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

Db 721 AGTATCTTGGACTGGACATCTTTAAACCAGGAGAGAAATGTTGATGTTGCTGACACGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280

Db 781 GGTGGAAATGACCCAGATATGTTAGTGTGGCACTTTGGCCTCAGCTGGAAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300

Db 841 GTAACCTGAGTGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320

Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340

Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGTGGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360

Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAACCGCAGAGAGTTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380

Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400

Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGAGAGTAGGGTCTTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420

Db 1201 TCAAGTTAAGAAGTACATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
QY 421 Met 421

Db 1261 ATG 1263

RESULT 6

US-10-103-327-13

; Sequence 13, Application US/10103327

; Patent No. 6890748
 ; GENERAL INFORMATION:
 ; APPLICANT: GARGER, Stephen A.
 ; APPLICANT: TURPEN, Thomas H.
 ; APPLICANT: KUMAGAI, Monto H.
 ; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
 ; FILE REFERENCE: 008010087CPUS06
 ; CURRENT FILING DATE: 2002-03-27
 ; PRIOR APPLICATION NUMBER: US/09/993,059
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 1284
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1) ... (1284)
 US-10-103-327-13

Alignment Scores:

Pred. No.: 8,25e-252 Length: 1284
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-10-103-327-13 (1-1284)

```

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTCTCTGGGACATCCCTGGGCTAGAGCACCTGGCAATGGATGGCAAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180
QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAAGGATGCAAGTTATAGTACTCTTCGATGTAGTACTGTGTGGTCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTGAGAAGGCAGACTTCAGCAGACCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 361 GCTAATATTATGTCACAGCAAGAGACTGAAGCTAGGAGTTATGTCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACTACGACATGTATGATGCCAGACCTTGTCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGACTGCTAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 540
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
DB 541 GCAGATGGTTATAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600
  
```

```

QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
DB 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTTCTGCATTCATGATTCCTGCGAAAGTATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGATGACCCAGATATGTTAGTGTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTTATTTCATGCTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGGAGACAACCTTTGAAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAAGCACCTCTCTCAGGCTTAGCCCTGGGCTAGCTATGATGATAAAACCGGACAGAGATTGGT 1080
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTCGCTCTTATACCATCGCATGCTTCCCTGGGTAAGGAGTGGCTGTAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
DB 1141 GCCTGTTTCATCACACAGCTCCTCCTGTGAAAGAGAGCTAGGGTTCATGATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTTACGTTAGTAAACACA 1260
QY 421 Met 421
DB 1261 ATG 1263
  
```

RESULT 7

US-09-993-059-3
 ; Sequence 3, Application US/09993059
 ; Patent No. 6887696
 ; GENERAL INFORMATION:
 ; APPLICANT: GARGER, Stephen A.
 ; APPLICANT: TURPEN, Thomas H.
 ; APPLICANT: KUMAGAI, Monto H.
 ; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
 ; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
 ; FILE REFERENCE: 008010087CPUS06
 ; CURRENT APPLICATION NUMBER: US/09/993,059
 ; CURRENT FILING DATE: 2001-11-13
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 1290
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1) ... (1290)
 US-09-993-059-3

Alignment Scores:

```
Pred. No.: 8,31e-252 Length: 1290
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-09-993-059-3 (1-1290)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGCATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAGGATGCAAGTATGAGTACCTCTGCATTGATGATGCTGTGGTGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAGAAGGCAGACCTTCAGGCAGACCTTCAGCGCTTCTCTCATGGGATTCGCCAGTA 360
QY 121 AlaAsnTrpValHisSerIysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 361 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGAGATTATGCAAGATGTTGGAATATA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGCTCCCTGGGAGTTTGGATACTAGCACTTGTGATGGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerIysLeuAsnLeu 180
DB 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGATTTGGAAAATTG 540
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
DB 541 GCAGATGGTTATAGCACATGCTTGGCCCTGGAATAGGACTGGCAGAGCAATTGTGTAC 600
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
DB 601 TCCTGTGAGTGCCCTCTTATATGTCGCCCTTCAAAGCCCAATATACAGAAATCGA 660
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAGTATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGATGACCCAGATATGTTAGTATGGCACTTTGGCTTCAGCTGGATCAGCA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACATCAGGCCCTCAAGCCAAAGCTCTCTCTTCAGGATAAGGAGCTAATTCCTCAAT 960
```

```
QY 321 GlnAspProLeuGlyLysGlnGlyTrpClnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGAGCAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAAACCGCAGGAGATTGGT 1080
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCCTGTAAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
DB 1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAAGGAGCTAGGCTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1201 TCAAGTTAAGAGTCACATTAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
QY 421 Met 421
DB 1261 ATG 1263

RESULT 8
US-10-103-327-3
; Sequence 3, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-10-103-327-3

Alignment Scores:
Pred. No.: 8,31e-252 Length: 1290
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-10-103-327-3 (1-1290)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
DB 121 ACCATGGGCTGGCTGCATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
```

Db 181 GATTCTCTGCATCAGTCAGAAAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCTCAGAAGGC 240
QY 81 TTPLYeAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrMetAlaProGlnArg 100
Db 241 TGAAGGATGCAAGGTATAGTACCTCTCATGTGATGACTGTGTGATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAAGGAGACTTCAGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerIysGlyLeuIysLeuGlyIleTyrAlaAspValGlyAsnIys 140
Db 361 GCTAATTATGTTACACAGCAAGGACTGAAGCTAGGCAATTTATGTCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCTCGGGAGTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
QY 161 AspTyrGlyValAspLeuIysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGACTGCTAATAATTTGATGTTGTACTGTGACAGTTTGGAAAATTG 540
QY 181 AlaAspGlyTyrIlyHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAGCATTTGTGTAC 600
QY 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnIysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGGCCCTCTTATATGTGGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrIysSerIleIys 240
Db 661 CAGTACTGCANCTACCTGGCGAAATTTTGTGTCATTTGATGATTCCTGGAAAAGTATTAAG 720
QY 241 SerIleLeuAspTyrThrSerPheAsnGlnIleArgIleValAspValalaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAACACGAGGAGAAATTTGTGATGTTGCTGGACAGGG 780
QY 261 GlyTyrAsnAspProAspMetIleValIleGlyAsnPheGlyLeuSerTyrPheGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTAGTGTGGCACTTTGGCCCTCAGCTGGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaIysAlaLeuLeuGlnAspIysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyIysGlnIleTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGCAAGCAAGGATACCAAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATTAACCCGCGAGAGATGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyIysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTCGGTAAAGGAGTGGCCGTGAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValIysArgLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValIleLeuGlnLeuGluAsnThr 420
Db 1201 TCAGGTTTAAGAAAGTACATTAATATCCACAGGCACGTGTTTGTCTTCAGCTAGAAAACACA 1260
QY 421 Met 421

Db 1261 ATG 1263
RESULT 9
US-09-993-059-9
; Sequence 9, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIWAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 08010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1296
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1296)
US-09-993-059-9
Alignment Scores:
Pred. No.: 8,37e-252 Length: 1296
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-10-602-219-12 (1-421) x US-09-993-059-9 (1-1296)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGACATCCCTGGGCTAGAGCCTGGCAATGGATTGGCAAGGACGCT 120
QY 41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCATCTGGAGGCTTCATGTCAACCTTGACTGCCAGGAAGGCCA 180
QY 61 AspSerCysIleSerGluIysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrrLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrMetAlaProGlnArg 100
Db 241 TGAAGGATGCAAGGTATGAGTACCTCTGCATTTGATGACTGTGTGATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAAGGAGACTTCAGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerIysGlyLeuIysLeuGlyIleTyrAlaAspValGlyAsnIys 140
Db 361 GCTAATTATGTTACACAGCAAGGACTGAAGCTAGGATTTATGTCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCTCGGGAGTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
QY 161 AspTyrGlyValAspLeuIysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGACTGCTAATAATTTGATGTTGTACTGTGACAGTTTGGAAAATTG 540
QY 181 AlaAspGlyTyrIysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200

Db 541 CGAGATGGTTAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCCTGTGAGTGGCCCTTTATATATGTCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCAATTTGATGATTCCTGGAAAGTATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAAATTTGTGATGTGCTGGACAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGATGATTTGGCACTTTGGCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGACAGGAGTACAGCTTAGACAGGAGACACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluLeGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGAGGAGATTGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGAGTGGCTGTAACTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAGAGAACTAGGGTTCTTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAGTCACATNAATCCACAGGCACTGTTTGTCTCAGCTAGAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263

RESULT 10

US-10-103-327-9

; Sequence 9, Application US/10103327

; Patent No. 6890748

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KOMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN

; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/103,327

; CURRENT FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 9

; LENGTH: 1296

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS
; LOCATION: (1) ... (1296)
US-10-103-327-9

Alignment Scores:

Pred. No.: 8,37e-252 Length: 1296
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-10-103-327-9 (1-1296)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATCAGCTGAGAAACCCAGAACTACATCTGGGCTGAGACACTGGCAATGGATTTGGCAAGGACGCT 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGACATCCCTGGGCTAGAGCACTGGCAATGGATTTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db 121 ACCATGGGCTGGTGCATCTGGGAGGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TCGAAGGATGACGGTTATGAGTACCTTGCAATTTGATGATGATGATGATGATGATGATGATGAT 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAAGGAGAGACTTCAGGACACCTTCAGCCCTTTCCTCATGGGATTCGACAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTCACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTCGCAGGCTTCCCTGGGAGTTTGGATACTACGACATTCATGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACATGTCCTTGGCCCTGATAGGACTGGCAGAGCAATGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCCTGTGAGTGGCCTCTTTATATATGTCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCAATTTGATGATTCCTGGAAAGTATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAAATTTGTGATGTGCTGGACAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900

QY 421 Met 421
DB 1261 ATG 1263

RESULT 12

US-10-103-327-5
; Sequence 5, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1308
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1308)
US-10-103-327-5

Alignment Scores:

Pred. No.: 8,49e-252 Length: 1308
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-10-103-327-5 (1-1308)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTGGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCTGGGACATCTCTGGGCTAGAGCACTGGACATGGATTGGCAAGAGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGTGCTGCCAGAGGCA 180
QY 61 AspSerCysIleSerGlnLeuLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAGGATGCAAGTTATGAGTACTCTGCAATGATGCTGTTGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTTCAGAAGGCAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTrpValHisSerIleGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 361 GCTAATTTATGTTTACAGCAAGAGACTGAAGCTAGAGATTATGCAAGATGTTGGAAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATCTACGACATTTGATGCTCCAGACACCTTTCCT 480

QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
DB 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGATAGACTGGCAGAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
DB 601 TCCTGTAGTGGCCTCTTATATGTGGCCCTTCAAAGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTTGATTCCTGGAAGATTAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGATTGTTGATGTTGCTGGACAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
DB 781 GGTGGAATGACCAGATATCTTGTAGTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGTGCTCTCTTTTATTCAATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACATATCAGCCTCAAGCCAAAGCTCTCTTCAAGATAGGAGCTTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGCGAAGCAAGGTTACCACTTAGACAGGAGAGCAACTTTGAACTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluLeuGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATATAAACCAGGAGAGATTGGT 1080
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACTGCTCTTATACCATCCAGCTGCTTCCCTGGTAAAGAGTGCCCTGTATATCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
DB 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAAGGAGCTAGGCTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1201 TCAAGGTTAAGAGTCACATAAATCCCAAGGCACTGTTTTGCTTTCAGCTAGAAAACACA 1260
QY 421 Met 421
DB 1261 ATG 1263

RESULT 13

US-08-928-881-18
; Sequence 18, Application US/08928881
; Patent No. 6083725
; GENERAL INFORMATION:
; APPLICANT: Selden et al., Richard F.
; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A
; TITLE OF INVENTION: DEFICIENCY
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/928,881
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fraser, Janis K.
 ; REGISTRATION NUMBER: 34,819
 ; REFERENCE/DOCKET NUMBER: 07236/003001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617/542-5070
 ; TELEFAX: 617/542-8906
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1343 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-928-881-18

Alignment Scores:

Pred. No.: 8,84e-252 Length: 1343
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-08-928-881-18 (1-1343)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 36 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 95
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 96 CTCGTTCTCGGACATCTCTGGGCTAGAGCACTGGACATGGATGGCAAGGAGCCCT 155
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 156 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 215
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 216 GATTCCTGCATCAGTGAGAAGCTTCTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 275
 QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 276 TGAAGGATGCAAGTTATGAGTACCTCTCATTTGATGACTGTGGATGGCTCCCAAGA 335
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 336 GATTCAGAGGCAGACTTCAGGCAGACCTTCAGCGCTTCTCTCATGGGATTCGCCAGCTA 395
 QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
 Db 396 GCTAATATTGTTACACAGCAAGGACTGAAGCTAGGAGTTATGCGATGTTGGAAATAA 455
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
 Db 456 ACTGGCGAGGCTTCTCTGGAGTTTGGATACTAGCATTTGATGCCAGACCTTTGCT 515
 QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 Db 516 GACTGGGAGTAGACTGCTAAATTTGATGGTTGTTACTGTGACAGCTTTGGAAATTTG 575
 QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
 Db 576 GCAGATGGTTAAGACATGCTCTGGCCCTGAATAGGACTGGCAGAGACATTTGTGTAC 635

QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
 Db 636 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 695
 QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 696 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTGGAAGATATAAG 755
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 756 AGTATCTTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTCGTGGACAGGG 815
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 816 GGTGGATGACCCAGATATGTTAGTGATTTGGCACTTTGGCCTCAGCTGGAATCAGCAA 875
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 876 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCTCTAAATGACCTC 935
 QY 301 ArgHisIleSerProGlnAlaIleValLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 936 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGAGTAAATTCCTCAATCAAT 995
 QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 996 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGGAGAGACAACTTTGAAGTGTG 1055
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1056 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGGCGAGGATTTGT 1115
 QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1116 GGACCTCGCTCTTATACCATCGAGTTCCTTCCTCGGGTAAAGGAGTGGCCTGTAACTCT 1175
 QY 381 AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
 Db 1176 GCCTGCTTCATCACACAGCTCTCTCTGTTGAAAGGAGTAGGTTCTTATGAATGGACT 1235
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 Db 1236 TCAAGGTTAAGAAGTACATAAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAATACA 1295
 QY 421 Met 421
 Db 1296 ATG 1298

RESULT 14

US-09-543-921-18
 ; Sequence 18, Application US/09543921
 ; Patent No. 6395884
 ; GENERAL INFORMATION:
 ; APPLICANT: Seldin et al., Richard F.
 ; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A DEFICIENCY
 ; NUMBER OF SEQUENCES: 28
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/543,921
 ; FILING DATE: 06-Apr-2000
 ; CLASSIFICATION: <Unknown>

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,881
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 07236/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1343 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-543-921-18

Alignment Scores:
Pred. No.: 8,846-252 Length: 1343
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-09-543-921-18 (1-1343)

QY 1 MetGlnLeuArgAspProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 36 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTCGCGCTTCCTGCGCC 95
QY 21 LeuValSerTrpAspIleProGlyValArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 96 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAGGAGCGCT 155
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 156 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGTGCGCAGGAGGCCA 215
QY 61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 216 GATTCCTGCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 275
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 276 TGGAAAGGATGCAGGTATGAGTACCTCTGCATGTGATGCTGTGTGATGGCTCCCAAGA 335
QY 101 AspSerGluGlyValGlnLeuAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 336 GATTCAGAAGGCAGACTTCAGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 395
QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 396 GCTAATATTATGTTACAGCAAGAGACTGAAGCTAGGAGTTTATGCGAGATGTTGGAATAA 455
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
DB 456 ACCTGCGCAGGCTTCCTCGGGAGTTTGGATACTAGCAATGTATGATGCCAGACCTTGTCT 515
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB 516 GACTGGGAGTGAAGCTGCTAAATTTGATGTTGTACTGTGACAGCTTGGAAATTTG 575
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
DB 576 GCAGATGGTTAAGACACATGCTCTGGCCCTGAATAGGACTGGCAGAGCAATTTGTGTAC 635
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
DB 636 TCCTGTGAGTGGCCTCTTTATATGATGTGGCCCTTTTCAAAAGCCCAATATATACAGAAATCGA 695

; GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
; CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATCTCGGAAAGTATAAG 755
; SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
; AGTATCTTGGACTGGACATCTTTTAAACGAGAGAGAAATTTGATGTTGCTGGACAGGG 815
; GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
; GGTGGAATGACCCAGATATGTTAGTATTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 875
; ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
; GTAATCTAGATGGCCCTCTGGGCTATCATGCTGCTCTTTATTCTATGCTTAATGACCTC 935
; ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
; CGACATCATGCCCTCAAGCCCAAGCTCTCTTCAGGATAGGAGTAAATGGCCATCAAT 995
; GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
; CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1055
; GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
; GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATATGATAAACCAGGAGATGGT 1115
; GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
; GGACCTCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCCTGTATCTCT 1175
; AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
; GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAGAGAGTAGGCTTCTATGATGAGCT 1235
; SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
; TCAAGGTTAAGAGTGCATATAATCCACAGGCACTGTTTGTCTTTCAGCTAGAAAATACA 1295
; Met 421
; ATG 1298
; Met 421
; ATG 1298

RESULT 15
; Sequence 3, Application US/09266014
; Patent No. 6458574
; GENERAL INFORMATION:
; APPLICANT: Selden, Richard F
; APPLICANT: Borowski, Marianne
; APPLICANT: Kinoshita, Carol M
; APPLICANT: Treco, Douglas A
; APPLICANT: Williams, Melanie D
; APPLICANT: Schuetz, Thomas J
; APPLICANT: Daniel, Peter F
; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency
; FILE REFERENCE: FABRY DISEASE (18082-001)
; CURRENT APPLICATION NUMBER: US/09/266,014
; CURRENT FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: 60/026,041
; PRIOR FILING DATE: 1996-09-13
; PRIOR APPLICATION NUMBER: 08/928,881
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: PCT US97/16603
; PRIOR FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-266-014-3
```

Alignment Scores:

Pred. No.: 8,84e-252 Length: 1343
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-09-266-014-3 (1-1343)

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 36 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 95
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 96 CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 155
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuApsCysGlnGluPro 60
DB 156 ACCATGGGCTGGCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 215
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 216 GATTCTCTGCATCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 275
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 276 TGGAGGATGCAGGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAGA 335
QY 101 AspSerGluGlyArgGluGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 336 GATTTCAGAGGAGAGACTTCAGGCAGACCCTCAGCGCTTCCCTCAGGGATTGCCAGCTA 395
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 396 GCTAATTATGTTACACAGCAAGGACTGAAGCTAGGGATTATGCAGATGTTGGAAATAA 455
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 456 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACTACGACATTCATGATGCCAGACCTTTGCT 515
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 516 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGATTGGAAAATTG 575
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 576 GCAGATGGTTATAGACACATGCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGTAC 635
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 636 TCCTGTGAGTGGCCTCTTTATATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCA 695
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 696 CAGTACTGCAATCACCTGGCGAAATTTGTGTCATTGATGATTCCTGGAAAAGTATAAG 755
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 756 AGTATCTTGGACTGGACATCTTTTACACAGAGAGAGAAATTTGTGATGTTGCTGGACAGGG 815
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 816 GGTGGAAATGACCCAGATATGTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 875
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 876 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 935
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
```

```
DB 936 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTGGCCATCAAT 995
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 996 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1055
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1056 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCCGACGAGATTGGT 1115
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1116 GGACCTCGCTCTTATATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCCTGTAACTCT 1175
QY 381 AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1176 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1235
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1236 TCAAGTTAAGAGTCACATTAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAATACA 1295
QY 421 Met 421
DB 1296 ATG 1298
```

Search completed: December 26, 2005, 13:52:44
Job time : 250 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 24, 2005, 08:46:48 ; Search time 230 Seconds
(without alignments)
1291.424 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MQRPELHLCALALRFLA.....RLRSHINPTGVLLQLENTM 421

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	429	1 AGAL_HUMAN	P06280 homo sapien
2	2314	100.0	429	2 Q53Y83_HUMAN	Q53Y83 homo sapien
3	2313	100.0	429	2 Q53HF3_HUMAN	Q53HF3 homo sapien
4	2178	94.1	403	2 Q6LER7_HUMAN	Q6LER7 homo sapien
5	1811	78.3	419	1 AGAL_MOUSE	P51569 mus musculus
6	1811	78.3	421	2 Q8GZ6_MOUSE	Q8GZ6 mus musculus
7	1480.5	64.0	429	2 Q90WL7_FURU	Q90WL7 fugu rubrip
8	1267	54.8	377	2 Q4RTE7_TETNG	Q4RTE7 tetraodon n
9	1130	48.8	338	2 Q5XJ76_BRARE	Q5XJ76 brachydanio
10	1074.5	46.4	431	2 Q6GQ17_XENLA	Q6GQ17 xenopus lae
11	1056.5	45.7	400	2 Q6GR44_XENLA	Q6GR44 xenopus lae
12	1051	45.4	405	1 NAGAB_CHICK	Q90744 gallus gall
13	1050	45.4	412	2 Q7Q1V0_ANOGA	Q7Q1V0 anopheles g
14	1038	44.9	410	2 Q4STX0_TETNG	Q4STX0 tetraodon n
15	1031	44.6	413	2 Q9VL27_DROME	Q9VL27 drosophila
16	1031	44.6	427	2 Q8MYX3_DROME	Q8MYX3 drosophila
17	1007.5	43.5	419	2 Q7Q6H3_ANOGA	Q7Q6H3 anopheles g
18	1006	43.5	417	2 Q9V7N9_DROME	Q9V7N9 drosophila
19	990.5	42.8	415	1 NAGAB_MOUSE	Q9QWR8 mus musculus
20	982	42.4	415	1 NAGAB_RAT	Q66H12 rattus norv
21	967.5	41.8	411	1 NAGAB_BOVIN	Q58DH9 bos taurus
22	952.5	41.2	411	1 NAGAB_HUMAN	P17050 homo sapien
23	862	37.3	451	2 Q21801_CAERL	Q21801 caenorhabdi
24	849	36.7	452	2 Q60LZ0_CAERL	Q60LZ0 caenorhabdi
25	784	32.6	399	2 Q9FWS4_SCHUA	Q9FWS4 schistosoma
26	729	31.5	413	2 Q5FWV8_ORYSA	Q5FWV8 oryza sativ
27	716.5	31.0	434	2 Q9LYL2_ARATH	Q9LYL2 arabidopsis
28	716.5	31.0	437	2 Q8VXZ7_ARATH	Q8VXZ7 arabidopsis
29	712	30.8	378	1 AGAL_COPAR	Q42656 coffea arab
30	710	30.7	466	2 Q54R54_DICDI	Q54R54 dictyostell
31	710	30.7	478	2 Q8VV86_CLOUO	Q8VV86 clostridium

RESULT 1

AGAL_HUMAN

ID AGAL_HUMAN STANDARD; PRT; 429 AA.

AC P06280;

DT 01-JAN-1988 (Rel. 06, Created)

DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)

DE Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-

DE galactoside galactohydrolase) (Alpha-D-galactosidase A) (Agalsidase

DE alfa)

GN Name-GLA;

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;

OC Homo.

OX NCBI_TaxID=9606;

RN [1]_TaxID=9606;

RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Lymphoblast;

RX MEDLINE=89263745; PubMed=2542896;

RA Kornreich R., Desnick R.J., Bishop D.F.;

RT "Nucleotide sequence of the human alpha-galactosidase A gene.;"

RL Nucleic Acids Res. 17:3301-3302(1989).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Fibroblast;

RX MEDLINE=87246603; PubMed=3036505;

RA Tsuji S., Martin B.M., Kaelow D.C., Migeon B.R., Choudary P.V.,

Stubblefield B.K., Mayor J.A., Murray G.J., Barranger J.A.,

Ginn S.E.;

RT "Signal sequence and DNA-mediated expression of human lysosomal alpha-

galactosidase A.;"

RL Eur. J. Biochem. 165:275-280(1987).

RN [3]

RP NUCLEOTIDE SEQUENCE.

RX MEDLINE=95352959; PubMed=7626884;

RA Oeltjen J.C., Liu X., Lu J., Allen R.C., Muzny D.M., Belmont J.W.,

Gibbs R.A.;

RT "Sixty-nine kilobases of contiguous human genomic sequence containing

the alpha-galactosidase A and Bruton's tyrosine kinase loci.;"

RL Mamm. Genome 6:334-338(1995).

RN [4]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RX PubMed=15772651; DOI=10.1038/nature03440;

RA Ross M.T., Grafham D.V., Coffey A.J., Scherer S., McLay K., Muzny D.,

Platzner M., Howell G.R., Burrows C., Bird C.P., Frankish A.,

Lovell F.L., Howe K.L., Ashurst J.L., Fulton R.S., Sudbrak R., Wen G.,

Jones M.C., Hattler M.E., Deadman R., Carter N.P., Hunt S.E.,

Ramsey J., Whittaker A., Deadman R., Scott C.E., Searle S.,

Crease A., Gunaratne P., Havlak P., Hodgson A., Metzker M.L.,

Richards S., Scott G., Steffen D., Sodergren E., Wheeler D.A.,

Worley K.C., Ainscough R., Ambrose K.D., Ansari-Lari M.A., Aradhya S.,

Ashevell R.I., Babbage A.K., Baggeley C.L., Ballabio A., Banerjee R.,

Barker G.E., Barlow K.F., Barrett I.P., Bates K.N., Beare D.M.,

Bell S.M., Bennett R., Bird C.P., Blyth P., Brown L., Brown R.,

Burns T., Burton J., Cawthon R., Chubb J., Church D., Clark R.,

Clee A., Clee D., Clee E., Clee J., Clee K., Clee L., Clee M., Clee N.,

Clee P., Clee R., Clee S., Clee T., Clee U., Clee V., Clee W., Clee X.,

Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D., Clee E., Clee F.,

Clee G., Clee H., Clee I., Clee J., Clee K., Clee L., Clee M., Clee N.,

Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T., Clee U., Clee V.,

Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D.,

Clee E., Clee F., Clee G., Clee H., Clee I., Clee J., Clee K., Clee L.,

Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T.,

Clee U., Clee V., Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B.,

Clee C., Clee D., Clee E., Clee F., Clee G., Clee H., Clee I., Clee J.,

Clee K., Clee L., Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R.,

Clee S., Clee T., Clee U., Clee V., Clee W., Clee X., Clee Y., Clee Z.,

Clee A., Clee B., Clee C., Clee D., Clee E., Clee F., Clee G., Clee H.,

Clee I., Clee J., Clee K., Clee L., Clee M., Clee N., Clee O., Clee P.,

Clee Q., Clee R., Clee S., Clee T., Clee U., Clee V., Clee W., Clee X.,

Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D., Clee E., Clee F.,

Clee G., Clee H., Clee I., Clee J., Clee K., Clee L., Clee M., Clee N.,

Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T., Clee U., Clee V.,

Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D.,

Clee E., Clee F., Clee G., Clee H., Clee I., Clee J., Clee K., Clee L.,

Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T.,

Clee U., Clee V., Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B.,

Clee C., Clee D., Clee E., Clee F., Clee G., Clee H., Clee I., Clee J.,

Clee K., Clee L., Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R.,

Clee S., Clee T., Clee U., Clee V., Clee W., Clee X., Clee Y., Clee Z.,

Clee A., Clee B., Clee C., Clee D., Clee E., Clee F., Clee G., Clee H.,

Clee I., Clee J., Clee K., Clee L., Clee M., Clee N., Clee O., Clee P.,

Clee Q., Clee R., Clee S., Clee T., Clee U., Clee V., Clee W., Clee X.,

Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D., Clee E., Clee F.,

Clee G., Clee H., Clee I., Clee J., Clee K., Clee L., Clee M., Clee N.,

Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T., Clee U., Clee V.,

Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D.,

Clee E., Clee F., Clee G., Clee H., Clee I., Clee J., Clee K., Clee L.,

Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T.,

Clee U., Clee V., Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B.,

Clee C., Clee D., Clee E., Clee F., Clee G., Clee H., Clee I., Clee J.,

Clee K., Clee L., Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R.,

Clee S., Clee T., Clee U., Clee V., Clee W., Clee X., Clee Y., Clee Z.,

Clee A., Clee B., Clee C., Clee D., Clee E., Clee F., Clee G., Clee H.,

Clee I., Clee J., Clee K., Clee L., Clee M., Clee N., Clee O., Clee P.,

Clee Q., Clee R., Clee S., Clee T., Clee U., Clee V., Clee W., Clee X.,

Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D., Clee E., Clee F.,

Clee G., Clee H., Clee I., Clee J., Clee K., Clee L., Clee M., Clee N.,

Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T., Clee U., Clee V.,

Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D.,

Clee E., Clee F., Clee G., Clee H., Clee I., Clee J., Clee K., Clee L.,

Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T.,

Clee U., Clee V., Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B.,

Clee C., Clee D., Clee E., Clee F., Clee G., Clee H., Clee I., Clee J.,

Clee K., Clee L., Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R.,

Clee S., Clee T., Clee U., Clee V., Clee W., Clee X., Clee Y., Clee Z.,

Clee A., Clee B., Clee C., Clee D., Clee E., Clee F., Clee G., Clee H.,

Clee I., Clee J., Clee K., Clee L., Clee M., Clee N., Clee O., Clee P.,

Clee Q., Clee R., Clee S., Clee T., Clee U., Clee V., Clee W., Clee X.,

Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D., Clee E., Clee F.,

Clee G., Clee H., Clee I., Clee J., Clee K., Clee L., Clee M., Clee N.,

Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T., Clee U., Clee V.,

Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D.,

Clee E., Clee F., Clee G., Clee H., Clee I., Clee J., Clee K., Clee L.,

Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T.,

Clee U., Clee V., Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B.,

Clee C., Clee D., Clee E., Clee F., Clee G., Clee H., Clee I., Clee J.,

Clee K., Clee L., Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R.,

Clee S., Clee T., Clee U., Clee V., Clee W., Clee X., Clee Y., Clee Z.,

Clee A., Clee B., Clee C., Clee D., Clee E., Clee F., Clee G., Clee H.,

Clee I., Clee J., Clee K., Clee L., Clee M., Clee N., Clee O., Clee P.,

Clee Q., Clee R., Clee S., Clee T., Clee U., Clee V., Clee W., Clee X.,

Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D., Clee E., Clee F.,

Clee G., Clee H., Clee I., Clee J., Clee K., Clee L., Clee M., Clee N.,

Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T., Clee U., Clee V.,

Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D.,

Clee E., Clee F., Clee G., Clee H., Clee I., Clee J., Clee K., Clee L.,

Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T.,

Clee U., Clee V., Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B.,

Clee C., Clee D., Clee E., Clee F., Clee G., Clee H., Clee I., Clee J.,

Clee K., Clee L., Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R.,

Clee S., Clee T., Clee U., Clee V., Clee W., Clee X., Clee Y., Clee Z.,

Clee A., Clee B., Clee C., Clee D., Clee E., Clee F., Clee G., Clee H.,

Clee I., Clee J., Clee K., Clee L., Clee M., Clee N., Clee O., Clee P.,

Clee Q., Clee R., Clee S., Clee T., Clee U., Clee V., Clee W., Clee X.,

Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D., Clee E., Clee F.,

Clee G., Clee H., Clee I., Clee J., Clee K., Clee L., Clee M., Clee N.,

Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T., Clee U., Clee V.,

Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B., Clee C., Clee D.,

Clee E., Clee F., Clee G., Clee H., Clee I., Clee J., Clee K., Clee L.,

Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R., Clee S., Clee T.,

Clee U., Clee V., Clee W., Clee X., Clee Y., Clee Z., Clee A., Clee B.,

Clee C., Clee D., Clee E., Clee F., Clee G., Clee H., Clee I., Clee J.,

Clee K., Clee L., Clee M., Clee N., Clee O., Clee P., Clee Q., Clee R.,

Clee S., Clee T.,

RA Beasley H., Beasley O., Beck A., Bethel G., Blechschmidt K., Brady N.,
 RA Bray-Allen S., Bridgeman A.M., Brown A.J., Brown M.J., Bonnin D.,
 RA Bruford E.A., Buhay C., Burch P., Burford D., Burgess J., Burrill W.,
 RA Burton J., Bye J.M., Carder C., Carrel L., Chako J., Chapman J.C.,
 RA Chavez D., Chen E., Chen G., Chen Y., Chen Z., Chinault C.,
 RA Ciccodicola A., Clark S.Y., Clarke G., Clee C.M., Clegg S.,
 RA Clerc-Blankenburg K., Clifford K., Cobley V., Cole C.G., Conquer J.S.,
 RA Corby N., Connor R.E., David R., Davies J., Davis C., Davis J.,
 RA Delgado O., Deshazo D., Dhani P., Ding Y., Dinh H., Dodsworth S.,
 RA Draper H., Dugan-Rocha S., Dunham A., Dunn M., Durbin K.J., Dutta I.,
 RA Eades T., Ellwood M., Emery-Cohen A., Erington H., Evans K.L.,
 RA Faulkner L., Francis F., Frankland J., Fraser A.E., Galgoczy P.,
 RA Gilbert J., Gill R., Gloeckner G., Gregory S.G., Gribble S.,
 RA Griffiths C., Grocock R., Gu Y., Gwilliam R., Hamilton C., Hart E.A.,
 RA Hawes A., Heath P.D., Heltman K., Hennig S., Hernandez J.,
 RA Hinzmann B., Ho S., Hoffs M., Howden P.J., Huckle E.J., Hume J.,
 RA Hunt P.J., Hunt A.R., Isherwood J., Jacob L., Johnson D., Jones S.,
 RA de Jong P.J., Joseph S.S., Keenan S., Kelly S., Kershaw J.K., Khan Z.,
 RA Kloschis P., Klages S., Knights A.J., Koslura A., Kovar-Smith C.,
 RA Laird G.K., Langford C., Lawlor S., Leversha M., Lewis L., Liu W.,
 RA Lloyd C., Lloyd D.M., Loulsegod H., Loveland J.E., Lovell J.D.,
 RA Lozado R., Lu J., Lyne R., Ma J., Maheshwari M., Matthews L.H.,
 RA McDowall J., McLaren S., McMurray A., Meidl P., Meitinger T.,
 RA Milne S., Miner G., Mistry S.L., Morgan A., Morris S., Mueller I.,
 RA Mullikin J.C., Nguyen N., Nordsiek G., Nyakatura G., O'dell C.N.,
 RA Okuwonu G., Palmer S., Pandian R., Parker D., Parrish J.,
 RA Perez L., Porter K.M., Ramsey Y., Reichwald K., Rhodes S.,
 RA Riddler K.A., Schlesinger D., Schueler M.G., Sehra H.K.,
 RA Shaw-Smith C., Shen H., Sheridan E.M., Showkeen R., Skuce C.D.,
 RA Smith M.B., Sotharan E.C., Steingrubber H.E., Steward C.A., Storey R.,
 RA Swann R.M., Swarbrick D., Taber P.E., Taudien S., Taylor T.,
 RA Teague B., Thomas K., Thorpe A., Timms K., Tracey A., Trevanion S.,
 RA Tromans A.C., d'Uro M., Verdusco D., Villaseana D., Waldron L.,
 RA Wall M., Wang Q., Warren J., Warry G.L., Wei X., West A.,
 RA Whitehead S.B., Whiteley M.N., Wilkinson J.E., Willey D.L.,
 RA Williams G., Williams L., Williamson A., Williamson H., Wilming L.,
 RA Woodmansey R.L., Wray P.W., Yen J., Zhang J., Zhou J., Zoghbi H.,
 RA Zorilla S., Buck D., Reinhardt R., Poustka A., Rosenthal A.F.,
 RA Lehrach H., Meindl A., Minx P.J., Hillier L.W., Willard H.F.,
 RA Wilson R.K., Waterston R.H., Rice C.M., Vaudin M., Coulson A.,
 RA Nelson D.L., Weinstock G., Sulston J.E., Durbin R., Hubbard T.,
 RA Gibbs R.A., Beck S., Rogers J., Bentley D.R.;
 RA "The DNA sequence of the human X chromosome.";
 RT Nature 434:325-337(2005).
 [5]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Uterus;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haie H.,
 RA Diatchenko L., Narusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bonak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smallus D.B.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [6]
 RP NUCLEOTIDE SEQUENCE OF 31-429, AND PARTIAL PROTEIN SEQUENCE.
 RC TISSUE=Lung;
 RX MEDLINE=86259694; PubMed=3014515;
 RA Bishop D.F., Calhoun D.H., Bernstein H.S., Hantzopoulos P., Quinn M.,
 RA Desnick R.J.;
 RT "Human alpha-galactosidase A: nucleotide sequence of a cDNA clone
 RT encoding the mature enzyme.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4859-4863(1986).
 [7]
 RN NUCLEOTIDE SEQUENCE OF 1-64
 RP MEDLINE=88112869; PubMed=2892762; DOI=10.1016/0378-1119(87)90374-X;
 RX Quinn M., Hantzopoulos P., Fidanza V., Calhoun D.H.;
 RA "A genomic clone containing the promoter for the gene encoding the
 RT human lysosomal enzyme, alpha-galactosidase A.";
 RL Gene 58:177-188(1987).
 [8]
 RN NUCLEOTIDE SEQUENCE OF 1-64
 RP MEDLINE=8834528; PubMed=2836863;
 RX Bishop D.F., Kornreich R., Desnick R.J.;
 RA "Structural organization of the human alpha-galactosidase A gene:
 RT further evidence for the absence of a 3' untranslated region.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:3903-3907(1988).
 [9]
 RN RNA EDITING OF POSITION 396.
 RP MEDLINE=95380278; PubMed=7503918;
 RX Novo F.J., Kruszewski A., McDermot K.D., Goldspink G., Gorecki D.C.;
 RA "Editing of human alpha-galactosidase RNA resulting in a pyrimidine to
 RT purine conversion.";
 RL Nucleic Acids Res. 23:2636-2640(1995).
 [10]
 RN X-RAY CRYSTALLOGRAPHY (3.45 ANGSTROMS) OF 32-422 IN COMPLEX WITH
 RP PRODUCT, HOMODIMERIZATION, AND N-GLYCOSYLATION.
 RX PubMed=15003450; DOI=10.1016/j.jmb.2004.01.035;
 RA Garman S.C., Garboczi D.N.;
 RT "The molecular defect leading to Fabry disease: structure of human
 RL alpha-galactosidase.";
 RL J. Mol. Biol. 337:319-335(2004).
 [11]
 RN REVIEW ON FD VARIANTS.
 RP MEDLINE=94258158; PubMed=7911050;
 RX Eng C.M., Desnick R.J.;
 RA "Molecular basis of Fabry disease: mutations and polymorphisms in the
 RT human alpha-galactosidase A gene.";
 RL Hum. Mutat. 3:103-111(1994).
 [12]
 RN VARIANT FD SER-40.
 RP MEDLINE=90092580; PubMed=2152885; DOI=10.1016/0014-5793(90)80046-L;
 RX Koide T., Ishiura M., Iwai K., Inoue M., Kaneda Y., Okada Y.,
 RA Uchida T.;
 RT "A case of Fabry's disease in a patient with no alpha-galactosidase A
 RT activity caused by a single amino acid substitution of Pro-40 by
 Ser.";
 RL FEBS Lett. 259:353-356(1990).
 [13]
 RN VARIANT FD VAL-296.
 RP MEDLINE=91101674; PubMed=1846223;
 RX von Scheidt W., Eng C.M., Fitzmaurice T.F., Erdmann E., Hubner G.,
 RA Olsen E.G.J., Christononou H., Kandolf R., Bishop D.F., Desnick R.J.;
 RT "An atypical variant of Fabry's disease with manifestations confined
 RT to the myocardium.";
 RL N. Engl. J. Med. 324:395-399(1991).
 [14]
 RN VARIANT FD GLN-301.
 RP MEDLINE=91022721; PubMed=2171331;
 RX Sakuraba H., Oshima A., Fukuhara Y., Shimamoto M., Nagao Y.,
 RA Bishop D.F., Desnick R.J., Suzuki Y.;
 RT "Identification of point mutations in the alpha-galactosidase A gene
 RT in classical and atypical hemizygotes with Fabry disease.";
 RL Am. J. Hum. Genet. 47:784-789(1990).
 [15]
 RN VARIANT FD TRP-356.
 RP MEDLINE=89198098; PubMed=2539398;
 RX Bernstein H.S., Bishop D.F., Astrin K.H., Kornreich R., Eng C.M.,
 RA Sakuraba H., Desnick R.J.;
 RT "Fabry disease: six gene rearrangements and an exonic point mutation
 RT in the alpha-galactosidase gene."

```
RL J. Clin. Invest. 83:1390-1399(1989).
RN [16]

Query Match      100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 2.8e-187;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPDLHGLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFPMCNLDQCEEP 60
DB 1 MOLRNPDLHGLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFPMCNLDQCEEP 60

QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDWNPQRDSGRLOADPQRPFGHGIQRL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDWNPQRDSGRLOADPQRPFGHGIQRL 120

QY 121 ANYVHSGKGLGIYADVGNKTCAGFPGSFGYYDIDAQTADWGVLLKFDGVCYCDSENL 180
DB 121 ANYVHSGKGLGIYADVGNKTCAGFPGSFGYYDIDAQTADWGVLLKFDGVCYCDSENL 180

QY 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFKPNYTEIRQYCNHWRNFADIDDSWKSJK 240
DB 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFKPNYTEIRQYCNHWRNFADIDDSWKSJK 240

QY 241 SILDWTSFNOERIVDVAGGGWNDPDMVLVGNFGLSWNOQVTOMALWAMAAPLFMSNDL 300
DB 241 SILDWTSFNOERIVDVAGGGWNDPDMVLVGNFGLSWNOQVTOMALWAMAAPLFMSNDL 300

QY 301 RHISPOAKALLQDKDVIAINQDPLGKQYQLRGDNFVWERPLSLGLAWAVAMINROEIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQYQLRGDNFVWERPLSLGLAWAVAMINROEIG 360

QY 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420

QY 421 M 421
DB 421 M 421

RESULT 2
Q53Y83 HUMAN
ID Q53Y83_HUMAN PRELIMINARY; PRT; 429 AA.
AC Q53Y83;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Galactosidase, alpha.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Maruyama K., Sugano S.;
RT "Oligo-capping: a simple method to replace the cap structure of
RT eucaryotic mRNAs with oligoribonucleotides.";
RL Gene 138:171-174 (1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Suzuki Y., Yoshitomo K., Maruyama K., Sugano A., Sugano S.;
RT "Construction and characterization of a full length-enriched and a 5'-
RT end-enriched cDNA library.";
RL Gene 200:149-156(1997).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
RA Tanaka A., Yokoyama S.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK222627; BAD96347.1; -; mRNA.
FT NON_TER 1
SQ SEQUENCE 429 AA; 48752 MW; 613P8BF21B107D7B CRC64;

Query Match      100.0%; Score 2314; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 2.8e-187;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPDLHGLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFPMCNLDQCEEP 60
DB 1 MOLRNPDLHGLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFPMCNLDQCEEP 60

QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDWNPQRDSGRLOADPQRPFGHGIQRL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDWNPQRDSGRLOADPQRPFGHGIQRL 120

QY 121 ANYVHSGKGLGIYADVGNKTCAGFPGSFGYYDIDAQTADWGVLLKFDGVCYCDSENL 180
DB 121 ANYVHSGKGLGIYADVGNKTCAGFPGSFGYYDIDAQTADWGVLLKFDGVCYCDSENL 180

QY 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFKPNYTEIRQYCNHWRNFADIDDSWKSJK 240
DB 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFKPNYTEIRQYCNHWRNFADIDDSWKSJK 240

QY 241 SILDWTSFNOERIVDVAGGGWNDPDMVLVGNFGLSWNOQVTOMALWAMAAPLFMSNDL 300
DB 241 SILDWTSFNOERIVDVAGGGWNDPDMVLVGNFGLSWNOQVTOMALWAMAAPLFMSNDL 300

QY 301 RHISPOAKALLQDKDVIAINQDPLGKQYQLRGDNFVWERPLSLGLAWAVAMINROEIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQYQLRGDNFVWERPLSLGLAWAVAMINROEIG 360

QY 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420

QY 421 M 421
DB 421 M 421

RESULT 3
Q53HF3 HUMAN
ID Q53HF3_HUMAN PRELIMINARY; PRT; 429 AA.
AC Q53HF3;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Galactosidase, alpha variant (fragment).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Maruyama K., Sugano S.;
RT "Oligo-capping: a simple method to replace the cap structure of
RT eucaryotic mRNAs with oligoribonucleotides.";
RL Gene 138:171-174 (1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Suzuki Y., Yoshitomo K., Maruyama K., Sugano A., Sugano S.;
RT "Construction and characterization of a full length-enriched and a 5'-
RT end-enriched cDNA library.";
RL Gene 200:149-156(1997).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
RA Tanaka A., Yokoyama S.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK222627; BAD96347.1; -; mRNA.
FT NON_TER 1
SQ SEQUENCE 429 AA; 48752 MW; 613P8BF21B107D7B CRC64;

Query Match      100.0%; Score 2313; DB 2; Length 429;
Best Local Similarity 99.8%; Pred. No. 3.4e-187;
Matches 420; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPDLHGLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFPMCNLDQCEEP 60
DB 1 MOLRNPDLHGLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFPMCNLDQCEEP 60
```

61	DSCTSEKLFMEAEMLVSEGWKADAGYEYLCIDDWMAPOQDSEGRLOADPQFPFGHRIOL	120	Db	1	PGARALDNGLARTPTMGWLHWFMCNLDCCQEPDSCISEKLFMEAEMLVSEGWKADAGY	60
61	DSCTSEKLFMEAEMLVSEGWKADAGYEYLCIDDWMAPOQDSEGRLOADPQFPFGHRIOL	120	QY	87	EYLCIDDDCWMAPOQDSEGRLOADPQFPFGHRIOLANYVHSKGLKLGIVADVGNKTCAGFP	146
121	ANYVHSKGLKLGIVADVGNKTCAGFPFGSGYDYDIDAQTFADGWVDLLKFDGVCDSLENL	180	Db	61	EYLCIDDDCWMAPOQDSEGRLOADPQFPFGHRIOLANYVHSKGLKLGIVADVGNKTCAGFP	120
121	ANYVHSKGLKLGIVADVGNKTCAGFPFGSGYDYDIDAQTFADGWVDLLKFDGVCDSLENL	180	QY	147	GSFGYDYDIDAQTFADGWVDLLKFDGVCDSLENLADGKYHMSLALNRTGRSIVYSCWPL	206
181	ADGYKMSLALNRTGRSIVYSCWPLMYPFQKPNYTEIROYCNHWRNFADIDDSWKSIX	240	Db	121	GSFGYDYDIDAQTFADGWVDLLKFDGVCDSLENLADGKYHMSLALNRTGRSIVYSCWPL	180
181	ADGYKMSLALNRTGRSIVYSCWPLMYPFQKPNYTEIROYCNHWRNFADIDDSWKSIX	240	QY	207	YMWPFQKPNYTEIROYCNHWRNFADIDDSWKSIXKILDTWTSFNQERIIVDVAGPGWINDPD	266
241	SILDWTSFNQERIIVDVAGPGWINDPDMLVIGNFGLSWNQVQTMALWAIMAAPLFMSNDL	300	Db	181	YMWPFQKPNYTEIROYCNHWRNFADIDDSWKSIXKILDTWTSFNQERIIVDVAGPGWINDPD	240
241	SILDWTSFNQERIIVDVAGPGWINDPDMLVIGNFGLSWNQVQTMALWAIMAAPLFMSNDL	300	QY	267	MLVIGNFGLSWNQVQTMALWAIMAAPLFMSNDLRHSPOAKALLQDKDVIAINQDPLGK	326
301	RHISPOAKALLQDKDVIAINQDPLGKQVQLRGDNFVWVERPLSLGLAWAVAMINROEIG	360	Db	241	MLVIGNFGLSWNQVQTMALWAIMAAPLFMSNDLRHSPOAKALLQDKDVIAINQDPLGK	300
301	RHISPOAKALLQDKDVIAINQDPLGKQVQLRGDNFVWVERPLSLGLAWAVAMINROEIG	360	QY	327	QGYQLRGDNFVWVERPLSLGLAWAVAMINROEIGGPRSYTIAVSLGKGVACNPACFIQ	386
361	GPRSYTIAVSLGKGVACNPACFITOLLPVKRLGFYEWTSRLRSHINPTGTVLLQLENT	420	Db	301	QGYQLRGDNFVWVERPLSLGLAWAVAMINROEIGGPRSYTIAVSLGKGVACNPACFIQ	360
361	GPRSYTIAVSLGKGVACNPACFITOLLPVKRLGFYEWTSRLRSHINPTGTVLLQLENT	420	QY	387	LLPVKRLGFYEWTSRLRSHINPTGTVLLQLENTM	421
421	M 421		Db	361	LLPVKRLGFYEWTSRLRSHINPTGTVLLQLENTM	395
421	M 421		RESULT 5			
Q6LER7_HUMAN						
ID	Q6LER7_HUMAN	PRELIMINARY;	PRT;	403	AA.	
AC	Q6LER7	2004 (TrEMBLrel. 27, Created)				
DT	05-JUL-2004	(TrEMBLrel. 27, Last sequence update)				
DT	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)				
DE	Alpha-galactosidase A (EC 3.2.1.22) (Fragment).					
GN	Name=alpha-Gala;					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;					
OC	Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	NUCLEOTIDE SEQUENCE.					
RC	TISSUE=Lung;					
RX	MEDLINE=86259694; PubMed=3014515;					
RA	Bishop D.F., Calhoun D.H., Bernstein H.S., Hantopoulos P., Quinn M.,					
RA	Desnick R.J.;					
RT	"Human alpha-galactosidase A: nucleotide sequence of a cDNA clone					
RT	encoding the mature enzyme."					
RL	Proc. Natl. Acad. Sci. U.S.A. 83:4859-4863(1986).					
RL	EMBL; D00039; BAA34059.1; -; mRNA.					
DR	SMR; Q6LER7; 6-396.					
DR	GO; GO:0004557; F:alpha-galactosidase activity; IEA.					
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.					
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.					
DR	InterPro; IPR002241; Glyco_hydro_27.					
DR	InterPro; IPR000111; Glyco_hydro_GHD.					
DR	Pfam; PF02065; Melibiase; 1.					
DR	PRINTS; PR00740; GLHYDRLASE27.					
DR	ProDom; PD002572; Glyco_hydro_GHD; 1.					
DR	PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.					
KW	Glycosidase; Hydrolase.					
FT	CHAIN	6	403	alpha-galactosidase A subunit peptide.		
FT	NON_TER	1	1			
SQ	SEQUENCE	403	AA;	45804	NW;	2FE193205BEB8D1A CRC64;
Query Match						
Best Local Similarity		94.1%	Score	2178;	DB 2;	Length 403;
Matches		395;	Conservative	0;	Mismatches	0;
					Indels	0;
					Gaps	0;
QY	27	PGARALDNGLARTPTMGWLHWFMCNLDCCQEPDSCISEKLFMEAEMLVSEGWKADAGY	86			

RA Klausner R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Altschul S.F., Zeeberg B., Wagner L., Shennan C.M., Schuler G.D.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uudin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahay J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences";
CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-
CC galactose residues in alpha-D-galactosides, including galactose
CC oligosaccharides, galactomannans and galactohydrolase.
CC -I- SUBUNIT: Homodimer.
CC -I- SUBCELLULAR LOCATION: Lysosomal.
CC -I- SIMILARITY: Belongs to the glycosyl hydrolase 27 family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC DR EMBL; U34071; AAA96749.1; -; mRNA.
CC DR EMBL; L46651; AAA74453.1; -; Genomic DNA.
CC DR EMBL; U58105; ABA7244.1; -; Genomic DNA.
CC DR EMBL; U50716; AAC52584.1; -; mRNA.
CC DR EMBL; U50715; AAC52583.1; -; Genomic DNA.
CC DR EMBL; BC09021; AAR09021.1; -; mRNA.
CC DR PIR; JC4522; JC4522.
CC DR SMR; P51569; 32-418.
CC DR Ensembl; ENSMUSG0000031266; Mus musculus.
CC DR MGI; MGI:1347344; Gla.
CC DR GO; GO:0005615; C:extracellular space; TAS.
CC DR InterPro; IPR002241; Glyco hydro 27.
CC DR InterPro; IPR000111; Glyco hydro_GHD.
CC DR Pfam; PF02065; Melibiose_1.
CC DR PRINTS; PR00740; GLHYDRLASE27.
CC DR ProDom; PD02572; Glyco_hydro_GHD; 1.
CC DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
CC KW Glycoprotein; Glycosidase; Hydrolase; Lysosome; Signal.
CC FT SIGNAL 1 31 By similarity.
CC FT CHAIN 32 419 Alpha-galactosidase A.
CC FT ACT_SITE 170 170 Nucleophile (By similarity).
CC FT ACT_SITE 231 231 Proton donor (By similarity).
CC FT SITE 203 207 Substrate binding (By similarity).
CC FT CARBOHYD 139 139 N-linked (GlcNAc...) (By similarity).
CC FT CARBOHYD 192 192 N-linked (GlcNAc...) (By similarity).
CC FT CARBOHYD 215 215 N-linked (GlcNAc...) (By similarity).
CC FT DISULFID 52 94 By similarity.
CC FT DISULFID 56 63 By similarity.
CC FT DISULFID 142 172 By similarity.
CC FT DISULFID 202 223 By similarity.
CC FT DISULFID 378 382 By similarity.
CC FT SEQUENCE 419 AA; 47643 MW; BD5E6A99AC113613 CRC64;
Query Match 78.3%; Score 1811; DB 1; Length 419;
Best Local Similarity 78.2%; Pred. No. 1.1e-144;
Matches 327; Conservative 41; Mismatches 50; Indels 0; Gaps 0;

QY 1 MQLRNPGLHCALRFLVSWDIPGARLDNGLARTPTWGLHWERFMCNLDQCEP 60
DB 1 MKLLSRDRLVCELALCPALVFWLSILGRALDNGLARTPTWGLHWERFMCNLDQCEP 60

QY 61 DSCISKLFPMWMAELMVSGWKDAGYEYICIDCWMAQORDSEGRLOADPQRPFGHRIQL 120
DB 61 DACISQLPQMAELMVSGWRDAGYDYLICIDCWMAQORDSEGRLOADPQRPFGHRIQL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGGYDIDAQTFADWGVLLKDGCCYCDLSLENL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGGYDIDAQTFADWGVLLKDGCCYCDLSVSL 180
QY 181 ADGYKMSLALNRTGRSIVYSCWPLMYMPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ENGKYMALALNRTGRSIVYSCWPLMYLPFPHKPNYTDIQQYCNHWRNFDDVYDSWESIK 240
QY 241 SILDWTSFQERIVDVAGPGGWNDDPMLVIGNFLGSWNOQVOTOMALWAIMAFLFMSNDL 300
DB 241 NILSWTVVYKQIEVAVAGPGGWNDDPMLVIGNFLGSWNOQVOTOMALWAIMAFLFMSNDL 300
QY 301 RHISPOAKALLQDKVIAINODPLGKQGYQLRQGDNFVWERPLSLGLAWAVAMINQEIG 360
DB 301 ROISSQAKALLQDKVIAINODPLGKQGYCFRKNHIEVWERPLSLGLAWAVAVRNQIEIG 360
QY 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEMTSRLRSHINPTGTVLLOLE 418
DB 361 GPCPYTIQISSLGRLACNPGCIIITQLPEKVLGFYEWTLTKTRVNSGTVLFRLE 418
RESULT 6
Q8BGZ6_MOUSE PRELIMINARY; PRT; 421 AA.
AC Q8BGZ6_MOUSE
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A430057F16 product:galactosidase, alpha, full insert
DE sequence (Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN
DE full-length enriched library, clone:E330039P08 product:galactosidase,
DE alpha, full insert sequence).
GN Name=Gla;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konegaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perte G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai K., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume M., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RL prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yanamoto R., Matsuoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Orawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RL sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai K., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi P., Takaku-Akaira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK040081; BAC30508.1; -; mRNA.
DR EMBL; AK054547; BAC35819.1; -; mRNA.
DR HSSP; P06280; 1R46.
DR SMR; QBGG26; 34-420.
DR MGI; MGI:1347344; Gta.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiose; 1.
DR PRINTS; PR00740; GLHYDLASE27.
DR ProDom; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA GALACTOSIDASE; 1.
DR SEQUENCE 421 AA; 47844 MW; 5F66772334014B6F CRC64;
SQ
Query Match 78.3%; Score 1811; DB 2; Length 421;
Best Local Similarity 78.2%; Pred. No. 1.1e-144;
Matches 327; Conservative 41; Mismatches 50; Indels 0; Gaps 0;
QY 1 MOLRNPELHLCALALRFLALYSWDIPGARALDNGLARTPTMGWLHWFRCNLDQCEEP 60
DB 3 MKLLSRDTRLVCELALCPALVFWSLGVRALDNGLARTPTMGWLHWFRCNLDQCEEP 62
QY 61 DSCISEKLFEMAEALMVSEGWKDAGYEYLCIDDCWMAPORDSEGRLOADPQRPPIRQL 120
DB 63 DACISQLFMQMAELMVSDGRDAGYDYLICIDDCWMAPERDSKGRLOADPQRPPIRQL 122
QY 121 ANYVHSKGLGIYADVGNKTCAGPGSGYDYDIDAGTADMGVDLLKPDGCGYCDLSLENL 180
DB 123 ANYVHSKGLGIYADVGNKTCAGPGSGYDYDIDAGTADMGVDLLKPDGCGYCDLSVSL 182
QY 181 ADGYSKMSLALNRTGRSIVVSCWPLYMPPFKPNTYETIRQYCNHWRNFADIDDSWKSIX 240
DB 183 ENGYKYNALNLTGRSIVVSCWPLYLAPFKPNTYDIQYCNHWRNFDDVYDSWESIK 242
QY 241 SILDWTSTFQERIVDVAGPGGWNDDMLVIGNFGLSNQOQVTOMALWAMAAPIFMSNDL 300
DB 243 NILSWTVVYQKEIVEVAGPGSWNDPMLVIGNFGLSNQOQVTOMALWAMAAPIFMSNDL 302
QY 301 RHISPOKALLQDKVDIATNODPLGKQYOLRGDNFEVWERPLSLGLAWAVAMINRQEIG 360
DB 303 RQISSQAKALLQNDVIATNODPLGKQYCFRKNHIEVWERPLSLGLAWAVAMINRQEIG 362
QY 361 GPRSYYTAVASLKGKGVACNPACFITOLLPVKRLGFGYEWTSRLRSHINPTGTVLLQLE 418
DB 363 GPCPYTIQISSLGRGLACNPGCIITOLLPEKHLGFGYEWTLTKTRVNPSTVLFRL 420
RESULT 7
Q90WL7_FUGRU PRELIMINARY; PRT; 429 AA.
AC Q90WL7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Alpha galactosidase a.
GN Name=agal;
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontoidea; Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Goode D., Elgar G.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Elgar G.S.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ290422; CAC44626.1; -; Genomic DNA.
DR HSSP; P06280; 1R46.
DR Ensembl; SINFRUG00000147523; Fugu rubripes.

Wed Dec 28 11:15:36 2005

```
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RA Director MGC Project;
RA Submitter (Oct-2004); to the EMBL/GenBank/DBJ databases.
DR EMBL; BC083209; AAH83209.1; -; mRNA.
DR ZFIN; ZDB-GENE-041010-207; zgc:101584.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR PRINTS; PR00740; GLHYDRLASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
SQ SEQUENCE 338 AA; 37846 MW; A3FC5C461B7C9649 CRC64;

Query Match 48.8%; Score 1130; DB 2; Length 338;
Best Local Similarity 60.3%; Pred. No. 4.1e-87;
Matches 210; Conservative 47; Mismatches 75; Indels 16; Gaps 2;

QY 68 LFMEMAEIYVSEGVKADAGYEYLCIDCWMAPOQDSEGRLOADPQRPFGHRIQLANYVHSK 127
DB 2 LDMSLCALMTA-GLRNKG-----THRGACRQTPKGFVPVASKLADYVHSK 45

QY 128 GLKLGIVADGNKTCAGFPSPGFGYDIDAQTFADWGVLDLKFDCYCDSENADGKHEM 187
DB 46 GLKLGIVADGNKTCAGFPSPGFGYDIDAQTFADWGVLDLKFDCYCDSENADGKHEM 105

QY 188 SLALNRTGRSIVSCWPLVMPKPNYTEIRQYCNHWRNFADIDDSWKSIXSLDWT 247
DB 106 SSALNRTGRSIVSCWPLVMPKPNYTEIRQYCNHWRNFADIDDSWKSIXSLDWT 165

QY 248 FNOQRIVDVAGPGWNPDMVLVGNFGLSNQVQVQVQVQVQVQVQVQVQVQVQVQVQV 307
DB 166 EKQIVVPVAGPGWNPDMVLVGNFGLSNQVQVQVQVQVQVQVQVQVQVQVQVQVQV 225

QY 308 KALLQDQVIAINQDPLGKQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQY 367
DB 226 KELLQNKQIIAINQDPLGKQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQY 285

QY 368 AVASLKGKGVACNPACFTIQLLPVKKLGFYEWTSRLSRSHINPTGTGVL 415
DB 286 SVAIMPSKWLCKNPKCNVQIILPTKEMGVQNLSEVVVQVNPVTGTTLL 333
```

RESULT 10

```
Q6GQ17 XENLA
ID Q6GQ17_XENLA PRELIMINARY; PRT; 431 AA.
AC Q6GQ17;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE LOC443592 protein (Fragment).
GN Name=LOC443592; (African clawed frog).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore J., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
```

```
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC072931; AAH72931.1; -; mRNA.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR PRINTS; PR00740; GLHYDRLASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
FT NON TER
SQ SEQUENCE 431 AA; 48787 MW; 4C578DB433983471 CRC64;

Query Match 46.4%; Score 1074.5; DB 2; Length 431;
Best Local Similarity 49.3%; Pred. No. 2.8e-82;
Matches 209; Conservative 62; Mismatches 126; Indels 27; Gaps 6;

QY 4 RNPELHGLCALRFLALVSWDIPGALDNGLARTTMCGLHWRFCNLDQCEPDSC 63
DB 18 RGYKMLTSLCLLTLMGLCW-----CLDNLGLVTPPMGMMTQWYRCNIDCKSDPNC 71

QY 64 ISEKLFMEAEIYVSEGVKADAGYEYLCIDCWMAPOQDSEGRLOADPQRPFGHRIQLANY 123
DB 72 ISENLIKIMADEMADSGWRDLGVYVICIDCWQKQKQKQKQKQKQKQKQKQKQKQKQ 131

QY 124 VHSKGLKGIYADVGNKTCAGFPSPGFGYDIDAQTFADWGVLDLKFDCYCDSENAD 182
DB 132 VHAKGLKGIYSDMGNTYTCGYPGTTLDTIKTDAETPASWEVDMLKFDGCVSNSTK-AL 190

QY 183 GYKMSLALNRTGRSIVSCWPLVMPKPNYTEIRQYCNHWRNFADIDDSWKSIXS 241
DB 191 GYKMSLALNRTGRSIVSCWPLVMPKPNYTEIRQYCNHWRNFADIDDSWKSIXS 250

QY 242 ILDTWTSFNOERIVDVAGPGWNPDMVLVGNFGLSNQVQVQVQVQVQVQVQVQVQVQV 301
DB 251 IIEWAKNQDVLQPAAGPGRWNPDMLTITGDFGLSYEQSKQLAIWALAPLMSNDLR 310

QY 302 HISPOAKALLQDQVIAINQDPLGKQYQYQYQYQYQYQYQYQYQYQYQYQYQY 361
DB 311 AISQAKDLLQNRLLIYINQSLKQGSLSIRGNLWVKEKRELINGQVYALAVLNKGTDL 370

QY 362 PRSYTIAVASLKGKGVACNPACFTIQLLP-----VKKLGFYEWTSRLSRSHINPTGT 413
DB 371 PRPYSTNLGLLN-----VTQCTDGYKLVNVEYKGLGMFTSTPIIRVNPTGT 420
```



```
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRLASE27.
DR ProDom; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
KW 3D-structure; Direct protein sequencing; Glycoprotein; Glycosidase;
KW Hydrolase; Lysozyme.
FT CHAIN 1 405 Alpha-N-acetylgalactosaminidase.
FT ACT_SITE 140 140 Nucleophile (By similarity).
FT ACT_SITE 201 201 Proton donor (By similarity).
FT CARBOHYD 161 161 N-linked (GlcNAc... ).
FT CARBOHYD 185 185 N-linked (GlcNAc... ).
FT CARBOHYD 369 369 N-linked (GlcNAc... ).
FT DISULFID 21 63
FT DISULFID 25 32
FT DISULFID 111 142
FT DISULFID 171 193
SQ SEQUENCE 405 AA; 45615 MW; E1EC0061739C305C CRC64;

Query Match 45.4%; Score 1051; DB 1; Length 405;
Best Local Similarity 53.0%; Pred. No. 2.5e-80;
Matches 206; Conservative 58; Mismatches 117; Indels 8; Gaps 6;

QY 32 LDNGLARTPTMGWLHWFRCNLDCCQEPDSCISEKLFMEAEIMVSEGKADAGEYLICI 91
DB 1 LENGLARTPPMGWLAWFRFCNVNCRBDRQCISEMLFMEMADRIAEDGWRELGYKYINI 60

QY 92 DDCWMAQRDSEGRQLADPQRFPHGIRQLANYVHSGKLGIVADVGNKTCAGFPG-SFG 150
DB 61 DDCWAAQRDAEGLVDPDFRPGIRKALADYVHARGLKLDIYDGLRLTCGGPGVTILD 120

QY 151 YYDIDAQTADMGVDLLKFGCGYCDLSLENLADGKMSLALNRTGRSIVYSCWPLYMWP 210
DB 121 RVEQDAQTFAEMGVDMKLDCGYSSGKEQ-AQGYPMARALNSTRPIVYSCSWPAFOGG 179

QY 211 F-QKPNTEIRQYCNHNRNFADIDDSWKSILDTWTSFNOERIVDVAGFGWNPDMVLV 269
DB 180 LPPKVNVTLLGEICNLWRNYDDIQDSWDSVLSIVDWFPTNQDVLQFAGEGHWNPDMLI 239

QY 270 IGNGLSNQOVTQMALWAIWAAPLMSNDRHISPOAKALLQDKVDVIAINQDPLGKQGY 329
DB 240 IGNGLSYEQSRQSMALWTMAAPLMSNDRHISPOAKALLQDKVDVIAINQDPLGKQGY 299

QY 330 Q-LRQGNFVYWERPLSGLAWAMINRQETGGPSRYTIAVASLGKGVACNPACFITQLL 388
DB 300 RIIEGSHIEVFLRPLSQASALVFFSRR-TDMFPRYTTSLAKLGFPMG---AAVEQDV 355

QY 389 PVKRLGFEYWTSLRSHINPTGVLLQL 417
DB 356 YSGKIISGLTKTDNFTVIINSPGVVMWYL 384
```

```
RESULT 13
Q7Q1V0 ANOGA
ID Q7Q1V0 ANOGA PRELIMINARY; PRT; 412 AA.
AC Q7Q1V0;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ENSANGP0000020847 (Fragment).
GN ORFNames=ENSANG00000018358;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
```

```
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
DR EMBL; AAAB01008980; EAA14548.2; -; Genomic DNA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRLASE27.
DR ProDom; PD002572; Glyco_hydro_GHD; 1.
FT NON_TER 1 1
FT NON_TER 412 412
SQ SEQUENCE 412 AA; 46517 MW; 4C91C8FB9EA6C88F CRC64;

Query Match 45.4%; Score 1050; DB 2; Length 412;
Best Local Similarity 49.0%; Pred. No. 3.1e-80;
Matches 193; Conservative 72; Mismatches 119; Indels 10; Gaps 4;

QY 31 ALDNGLARTPTMGWLHWFRCNLDCCQEPDSCISEKLFMEAEIMVSEGKADAGEYLIC 90
DB 17 SLENGLARTPPMGWLAWFRFCNVNCRBDRQCISEMLFMEMADRIAEDGWRELGYKYINI 76

QY 91 IDCWMAQRDSEGRQLADPQRFPHGIRQLANYVHSGKLGIVADVGNKTCAGFPG-SFG 150
DB 77 VDCWLEKSGRPGELVADRRRFPSCWKALANVYHAKLGFIVDGYNTCAGYPGILG 136

QY 151 YYDIDAQTADMGVDLLKFGCGYCDLSLENLADGKMSLALNRTGRSIVYSCWPLY-MW 209
DB 137 FSANDAQAQFASWVDVYVLDGCGYSLPID-MDHGYPEFRNLNATGRPMIYSCSWPVQIY 195

QY 210 PFQKPNTEIRQYCNHNRNFADIDDSWKSILDTWTSFNOERIVDVAGFGWNPDMVLV 269
DB 196 AGNPNYSSIIQCNLWRNYDDIQDSWASLESIIYGNQDAIIPNAGEHWNPDMLI 255

QY 270 IGNGLSNQOVTQMALWAIWAAPLMSNDRHISPOAKALLQDKVDVIAINQDPLGKQGY 329
DB 256 IGNGLSYEQSRQSMALWAIWAAPLMSNDRHISPOAKALLQDKVDVIAINQDPLGKQGY 315

QY 330 QLRQGNFVYWERPLSGL-----AWAVAMINRQETGGPSRYTIAVASLGKGVACNPACFI 384
DB 316 RIYKHGIEIWSRPITPIYQTYYSIAFAFYNRRTDGTSPDVAVTLRGLG---LISPTGYR 372

QY 385 TQLLPVKRLGFEYWTSLRSHINPTGVLLQL 418
DB 373 VEDLYEEVDYGLSPQTKIKVKNPSPGVILRAD 406
```

```
RESULT 14
Q4STX0 TETNG
ID Q4STX0 TETNG PRELIMINARY; PRT; 410 AA.
AC Q4STX0;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF14084, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00012743001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontoidea; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Bouneau L., Fischer C., Ozouf-Costaz E., Bernot A.,
Mauceli E., Brunet F., Petit J.L., Stange-Thomann N.,
Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
```

RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castella M., Katinka M., Vacherie B.,
RA Blomont C., Skalli Z., Cattolico L., Poullain J., De Bernardis V.,
RA Cruaud C., Duprat S., Bottier P., Coutanceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.,
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CA01014084; CAP95912.1; -; Genomic_DNA.
FT NON TER 410 410
SQ SEQUENCE 410 AA; 46172 MW; F5C661DA72DCB88 CRC64;

Query Match 44.9%; Score 1039; DB 2; Length 410;
Best Local Similarity 48.9%; Pred. No. 3.2e-79;
Matches 207; Conservative 58; Mismatches 116; Indels 42; Gaps 8;

QY 8 LHLGALALRFLALVSWDIPGARALDGLARTPTMGWLHWRMCMNLDQCEPDCISEK 67
DB 1 MHLALLLLASVLSL-----GTALLDGLMRTPTMGWLHWRMCMNLDQCEPDCISEK 54

QY 68 LFMEMALMVSEGWDAGYEYLCIDCWAPQDSRGLQADQRPFGIRQLAVYVHSK 127
DB 55 LFTDMADRLSQDCKWGLGYVYVNIIDCWSSKDKKGRQLQDPKRPFGGPKLARYMHDR 114

QY 128 GLKGLIYADVGNKTCAGFGPS-PGYVDIDAQTADNGVLLKFDGCGYCSLENLADGYKH 186
DB 115 GLKGLIYDGMGLTTCGGYFGTGLDKIDAKTAEWEVDMLKFDGCGYSNEVEQ-QQGYPL 173

QY 187 MSALNRTGSRIVYSCWPLVMPFP-QKPNYTIROYCNHWRNFADIDDSWKSILQW 245
DB 174 MSALNATGRPIAYSCWPAYSGLPLPKVNYTQGLICNLWRNYDDIQDSWSDVNLIDW 233

QY 246 TSFNQERIVDVAGFGWNPDMVLVIGNFGLSMNQVTOALWALMAAPLFMSNDLRHISP 305
DB 234 FPDNQDVIAAPAGFGWRNPDMVLVIGFGLSLQSRTOALWALMAAPLFMSNDLRHISA 293

QY 306 QAKALLQDQVIAINQDPLGKQYQL-RQGDNFEVWERPISGLAWAVINRQETGGPRS 364
DB 294 EARSILQNKIAIGINQDPLGFGQRLVYKESGIEVFWRLSNDTSALVFFSRR----- 346

QY 365 YTIASVSLGKGVACNPFITQLLPVKRLGYE-----WTSRLRS-----HINPTG 411
DB 347 -----TDMPIRYKTSLSKUSYPTGYTKYVDVFTQKSSLKDSADFVSVNPTG 394

QY 412 TVL 414
DB 395 VVM 397

RESULT 15
Q9VL27 DROME PRELIMINARY; PRT; 413 AA.
AC Q9VL27;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CG5731-PA.
GN Names=CG5731; ORFNames=CG5731;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananthanathan P.G., Scherer S.E., Li P.W., Hoskins R.A., Calle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Galbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hosain D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.L., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Weasman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RT Science 287:2185-2195(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective";
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.B.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Galbart W.M., Rubin G.M.,

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 26, 2005, 10:28:11 ; Search time 5498 Seconds
(without alignments)
4352.691 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MQLRNPEHLGCALALRFLA.....RLRSHNPTGTVLQLENTM 421

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO spool_p/US10602219/runat_23122005_151141_17136/app.query.fasta_1.583
-DB=GenEmbl -QFMT=fastcap -SURFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10602219@cgn 1.7415 @runat_23122005_151141_17136 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.in.*
3: gb.env.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pr.*
9: gb.ro.*
10: gb.sts.*
11: gb.sy.*
12: gb.un.*
13: gb.vl.*
14: gb.hcg.*
15: gb.pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	1266	6	AR653696 Sequence
2	2314	100.0	1266	6	AR654964 Sequence
3	2314	100.0	1278	6	AR653694 Sequence

4	2314	100.0	1278	6	AR654962	Sequence
5	2314	100.0	1284	6	AR653697	Sequence
6	2314	100.0	1284	6	AR654965	Sequence
7	2314	100.0	1290	6	CS135670	Sequence
8	2314	100.0	1290	6	AR653692	Sequence
9	2314	100.0	1290	6	AR654960	Sequence
10	2314	100.0	1290	8	BT006864	Homo sapi
11	2314	100.0	1290	11	AY889631	Synthetic
12	2314	100.0	1290	11	AY889632	Synthetic
13	2314	100.0	1290	11	AY892090	Synthetic
14	2314	100.0	1290	11	BT007835	Synthetic
15	2314	100.0	1296	6	AR653695	Sequence
16	2314	100.0	1296	6	AR654963	Sequence
17	2314	100.0	1306	6	BD182986	Pharmaceu
18	2314	100.0	1306	6	BD185988	Pharmaceu
19	2314	100.0	1308	6	AR653693	Sequence
20	2314	100.0	1308	6	AR654961	Sequence
21	2314	100.0	1312	8	BC002689	Homo sapi
22	2314	100.0	1319	6	AX840234	Sequence
23	2314	100.0	1319	8	HSAGALAR	Human mRNA
24	2314	100.0	1343	6	AR101984	Sequence
25	2314	100.0	1343	6	BD064229	Therapy f
26	2314	100.0	1343	6	BD247365	Treatment
27	2314	100.0	1343	6	CS105027	Sequence
28	2314	100.0	1343	6	AR234439	Sequence
29	2314	100.0	1343	6	AR322462	Sequence
30	2314	100.0	1343	6	AR372138	Sequence
31	2314	100.0	1350	6	CQ899396	Sequence
32	2314	100.0	1350	6	CS033394	Sequence
33	2314	100.0	1350	6	CS035758	Sequence
34	2314	100.0	1350	6	CS042346	Sequence
35	2314	100.0	1350	6	CS044710	Sequence
36	2314	100.0	1352	6	AX523899	Sequence
37	2314	100.0	1390	6	AX780273	Sequence
38	2314	100.0	1393	6	I30313	Sequence
39	2314	100.0	1393	6	AX030859	Sequence
40	2314	100.0	1447	6	BD134457	Human nuc
41	2314	100.0	1447	6	AX017286	Sequence
42	2314	100.0	1447	6	AX524990	Sequence
43	2314	100.0	3809	6	AX282869	Sequence
44	2313	100.0	1341	8	AX222627	Homo sapi
45	2310	99.8	1350	6	CQ882046	Sequence

ALIGNMENTS

RESULT 1	AR653696	Sequence 11	1266 bp	DNA	linear	PAT 13-JUN-2005
LOCUS	AR653696	Sequence 11	from patent US 6887696.			
DEFINITION	AR653696					
ACCESSION	AR653696					
VERSION	AR653696.1	GI:67584281				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 1266)					
AUTHORS	Gargier, S.J., Turpen, T.H. and Kumagai, M.H.					
TITLE	Production of lysosomal enzymes in plants by transient expression					
JOURNAL	Patent: US 6887696-A 11 03-MAY-2005;					
FEATURES	Large Scale Biology Corporation; Vacaville, CA					
source	Location/Qualifiers					
	1..1266					
	/organism="unknown"					
	/mol_type="genomic DNA"					
ORIGIN						
Alignment Scores:						
Pred. No.:	1.34e-207	Length:	1266			
Score:	2314.00	Matches:	421			
Percent Similarity:	100.00%	Conservative:	0			
Best Local Similarity:	100.00%	Mismatches:	0			
Query Match:	100.00%	Indels:	0			

DB:	6	Gaps:	0
US-10-602-219-12 (1-421) x AR653696 (1-1266)			
QY	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
DB	1	ATGCAGCTGAGGAACCCAGAACTTACATCTGGGCTGGCGCTTGGCTTCCTGGCC	60
QY	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
DB	61	CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATGGCAAGGCGCT	120
QY	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
DB	121	ACCATGGGCTGGCTGGACCTGGGAGCGCTTCATGTGCACCTTGACTGCCAGGAAGCCA	180
QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
DB	181	GATTCCTGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC	240
QY	81	TrpLysAspAlaGlyTrpGluArgPheMetCysAsnLeuAspCysTrpMetAlaProGlnArg	100
DB	241	TGGAAGGATGCAAGGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC	300
QY	321	GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp	340
DB	961	CAGGACCCCTTGGGCAAGGAGGTACCGCTTAGACAGGAGACAACTTCAAGTGTGG	1020
QY	341	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly	360
DB	1021	GBACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCAGGAGATTGCT	1080
QY	361	GlyProArgSerTyThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro	380
DB	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAACTCT	1140
QY	381	AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTyArgGluTrpThr	400
DB	1141	GCCTGCTTCATCACAGCTCCTCCTGTGAAAGAAAGCTAGGTTTCTATGAATGGACT	1200
QY	401	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr	420
DB	1201	TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAACACA	1260
QY	421	Met 421	
DB	1261	ATG 1263	
RESULT 2			
AR654964			
LOCUS			
Sequence 11 from patent US 6890748.			
DEFINITION			
AR654964			
ACCESSION			
AR654964.1 GI:67586503			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Unknown.			
Unclassified.			
REFERENCE			
1 (bases 1 to 1266)			
AUTHORS			
Garger, S.J., Turpen, T.H. and Kumagai, M.H.			
TITLE			
Production of lysosomal enzymes in plants by transient expression			
JOURNAL			
Patent: US 6890748-A 11 10-MAY-2005;			
Large Scale Biology Corporation; Vacaville, CA			
FEATURES			
Location/Qualifiers			
1..1266			
/organism="unknown"			
/mol_type="genomic DNA"			
source			
ORIGIN			
Alignment Scores:			
Pred. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			
DB:			
US-10-602-219-12 (1-421) x AR654964 (1-1266)			
QY	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
DB	1	ATGCAGCTGAGGAACCCAGAACTTACATCTGGGCTGGCGCTTGGCTTCCTGGCC	60
QY	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
DB	61	CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATGGCAAGGCGCT	120
QY	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
DB	121	ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCACCTTGACTGCCAGGAAGCCA	180
QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
DB	181	GATTCCTGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC	240
QY	81	TrpLysAspAlaGlyTrpGluArgPheMetCysIleAspAspCysTrpMetAlaProGlnArg	100
DB	241	TGGAAGGATGCAAGGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC	300
QY	101	AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
DB	301	GATTCAGAAGGCGAGCTTCAGGAGACCCCTCAGCGCTTTCCTCATGGGATTGCCAGCTA	360

QY 121 AlaAspTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 DB 361 GCTAATTATGTTACAGCAAGAGCTGAAGCTAGGATTTATGACAGATGTTGGAAATAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 DB 421 ACCTGCGAGGCTCCCTGGGAGTTTGGATACACGACATTTGATGACAGCTTTGCT 480
 QY 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 DB 481 GACTGGGAGTAGATCTCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
 QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAlaLeuAlaLeuAlaLeuAlaLeu 200
 DB 541 GCAGATGCTTATAAGCACATGCTTGGCCCTGAAATAGACTGGCAGAAAGCATTTGTAT 600
 QY 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
 DB 601 TCCTGTAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
 DB 661 CAGTACTGCATCACTGCGCAATTTTGTGACATTTGATGCTTCTGGAAAGTATAAG 720
 QY 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 721 AGTATCTTGGACTGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
 QY 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGln 280
 DB 781 GGTGGAAATGACCCAGATATGTTAGTGTATGGCACTTTGGCCCTCAGCTGGAATCAGAA 840
 QY 281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTTAAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATGGCCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
 DB 961 CAGGACCCCTTGGGCAAGCAAGGTACACAGCTTAGACAGGGAGACAACCTTGAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
 DB 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGCTGTAGCTATGATTAACCGGAGGAGATGGT 1080
 QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 DB 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCCCTGTAATCCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
 DB 1141 GCCTGCTTTCATCACAGCTCTCCCTGTGAAAGAAAGCTAGGCTTCTATGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 DB 1201 TCAAGTTAAGAAGTACATAAATCCACAGGCACTGTTTTCTTCAGCTAGTAAACACA 1260
 QY 421 Met 421
 DB 1261 ATG 1263

RESULT 3
 LOCUS AR653694
 DEFINITION Sequence 7 from patent US 6887696.
 ACCESSION AR653694
 VERSION AR653694.1 GI: 67584279
 KEYWORDS Unknown.

ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1278)
 AUTHORS Garger, S.J., Turpen, T.H. and Kumagai, M.H.
 TITLE Production of lysosomal enzymes in plants by transient expression
 JOURNAL Patent: US 6887696-A 7 03-MAY-2005;
 FEATURES Large Scale Biology Corporation; Vacaville, CA
 source location/Qualifiers
 1..1278
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.35e-207 Length: 1278
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-602-219-12 (1-421) x AR653694 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCCTTCCTCTGGCC 60
 QY 21 LeuValSerTyrAspIleProGlyValAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 DB 61 CTGCTTCTGGGACATCCCTGGGCTAGACACTGGACAATGGATTGGCAAGGACGCT 120
 QY 41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 DB 121 ACCATGGGCTGGCTGCACTGGGAGGCTTCATGTGCAACCTTGACTGCGAGAGAGCCA 180
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 DB 181 GATTCCTGCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGCC 240
 QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 DB 241 TGGAAAGGATGAGGTTATGAGTACTCTTCGATTCATGATGATGCTGGTGGCTCCCAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 DB 301 GATTGAGAAGGAGAGCTTCAGGACACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
 QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 DB 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTTAGGATTTATGACAGATGTTGGAATAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 DB 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATACACGACATTTGATGCGCCAGACCTTGTCT 480
 QY 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 DB 481 GACTGGGAGTAGATCTCTAAATTTGATGGTTGTTACTGTGACAGCTTTGGAAATTTG 540
 QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 DB 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAAATAGGACTGGCAGAAAGCATTTGTAT 600
 QY 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
 DB 601 TCCTGTAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
 DB 661 CAGTACTGCATCACTGCGCAATTTTGTGACATTTGATGCTTCTGGAAAGTATAAG 720
 QY 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260


```

Db      1141 GCCTGCTTCATCACACAGCTCCTCCCTGTAAGGAAGAGCTAGGCTTCTATGAATGGACT 1200
Qy      401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db      1201 TCAAGGTTAAGAGTCACATAAATCCACAGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
Qy      421 Met 421
Db      1261 ATG 1263

RESULT 5
AR653697
LOCUS      AR653697      1284 bp      DNA      linear      PAT 13-JUN-2005
DEFINITION Sequence 13 from patent US 6887696.
ACCESSION AR653697
VERSION    AR653697.1 GI:67584282
KEYWORDS   Unknown.
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 1284)
AUTHORS     Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE        Production of lysosomal enzymes in plants by transient expression
JOURNAL      Patent: US 6887696-A 13 03-MAY-2005;
              Large Scale Biology Corporation; Vacaville, CA
FEATURES
  source
    1..1284
    /organism="unknown"
    /mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.:      1,36e-207      Length:      1284
Score:          2314.00      Matches:     421
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:              6      Gaps:         0

US-10-602-219-12 (1-421) x AR653697 (1-1284)
Qy      1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db      1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
Qy      21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db      61 CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120
Qy      41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db      121 ACCATGGGCTGGCTGCATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
Qy      61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db      191 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy      81 TrpLysAspAlaGlyTyTrpGluTyTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db      241 TGGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
Qy      101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db      301 GATTCAGAAGGCAGACTTCAGCAACCTTCAGCCCTTTCCTCATGGGATTCGCCAGCTA 360
Qy      121 AlaAsnTyTrpValHisSerLysGlyLeuLysLeuGlyIleTyTrpAlaAspValGlyAsnLys 140
Db      361 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
Qy      141 ThrCysAlaGlyPheProGlySerPheGlyTyTrpAspIleAspAlaGlnThrPheAla 160
Db      421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACTACGACATTGATGCGCCAGACCTTTGCT 480

```

```

Qy      161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyTrpCysAspSerLeuGluAsnLeu 180
Db      481 GACTGGGAGGTAGATCTGCTAAATTTTGGTTTGTACTGTGACAGCTTTCGAAAAATTTG 540
Qy      181 AlaAspGlyTyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTy 200
Db      541 GCAGATGGTTATATAGCACATGCTCTGGCCCTGTAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Qy      201 SerCysGluTrpProLeuTyTrpMetTrpProPheGlnLysProAsnTyTrpGluIleArg 220
Db      601 TCCGTGAGTGGCCTCTTTATATGCTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy      221 GlnTyTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db      661 CAGTACTGCAATCACTGCGCAAAATTTTGTGCTGACATTGATGATCTCTGGAAGATATAAG 720
Qy      241 SerIleLeuAspTrpThrSerPheAsnGlnLysArgIleValAspValAlaGlyProGly 260
Db      721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGGACCAAGG 780
Qy      261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db      781 GGTGGAAATGCCAGATATGTTAGTGTATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy      281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db      841 GTAACACTCAGATGGCCCTCTGGGCTATCATGCTGCTCCTTTTATTTCATGCTTAATGACCTC 900
Qy      301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGCTTAATTGCCATCAAT 960
Qy      321 GlnAspProLeuGlyLysGlnGlyTyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db      961 CAGGACCCCTTGGCAAGCAAGGGTACCCAGCTTAGACAGGAGAGACAATTTGAAGTGTGG 1020
Qy      341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTAGCTATGATAAACCAGCAGAGATTGCT 1080
Qy      361 GlyProArgSerTyTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1081 GGACCTCGCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Qy      381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyTrpGluTrpThr 400
Db      1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGTTCTATGAATGGACT 1200
Qy      401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db      1201 TCAAGGTTAAGAGTCACATAAATCCACAGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
Qy      421 Met 421
Db      1261 ATG 1263

RESULT 6
AR654965
LOCUS      AR654965      1284 bp      DNA      linear      PAT 13-JUN-2005
DEFINITION Sequence 13 from patent US 6890748.
ACCESSION AR654965
VERSION    AR654965.1 GI:67586504
KEYWORDS   Unknown.
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 1284)
AUTHORS     Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE        Production of lysosomal enzymes in plants by transient expression
JOURNAL      Patent: US 6890748-A 13 10-MAY-2005;
              Large Scale Biology Corporation; Vacaville, CA
FEATURES
  Location/Qualifiers

```


Db 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTTCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGAGGTTATAGTACTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAAGGCAGACTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGGACTGAAGCTAGGAGTATATGCAAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGGAGGCTTCCCTGGGAGTTTGGATACTAGCAATGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTGTACTGTGACAGATTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTATAGACACATGTCCTTGGCCCTGTAATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTCTGCATTTGATGATTCCTGGAAAGTATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnAlaArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGACATCTTTAAACCAGGAGAGAAATTTGTATGTTGCTGGACAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGAGGGTACCAGCTTTAGACAGGGAGACAATTTGAAAGTGGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCCGACAGGAGTTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTGTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTTCATCACAGCTCTCCCTGTGAAAGGAGGAGCTAGGGTTCTATGAATGGACT 1200

Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAGTGCATATAATCCACAGGCACCTGTTTGTCTTCAGCTAGAAAATACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
RESULT 8
LOCUS AR653692 1290 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 3 from patent US 6887696.
ACCESSION AR653692
VERSION AR653692.1 GI:67584277
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 3 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source location/Qualifiers
1..1290
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1.37e-207 Length: 1290
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-602-219-12 (1-421) x AR653692 (1-1290)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGACGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCTCGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTTCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGAGGTTATAGTACTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAAGGCAGACTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGGACTGAAGCTAGGAGTATATGCAAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGGAGGCTTCCCTGGGAGTTTGGATACTAGCAATGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180

Db	481	GACTGGGAGTAGATCTGTAATAATTTGATGCTTGTCTGTGACAGATTTGGAAATTTG	540
Qy	181	AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr	200
Db	541	GCAGATGGTTATTAAGCACATGTCCTTGGCCCTGAATAGACTGGCAGAACATTTGTGTAC	600
Qy	201	SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg	220
Db	601	TCCTGTGAGTGGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA	660
Qy	221	GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys	240
Db	661	CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATAAAG	720
Qy	241	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly	260
Db	721	AGTATCTTGGACTGGACATCTTTTAACCCAGGAGAGAAATTTGTGTGCTGGACCAAGG	780
Qy	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln	280
Db	781	GGTTGGAATGACCCAGATATGTAGTATGGCAACTTTTGGCCTCAGCTGGAAATCAGCAA	840
Qy	281	ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu	300
Db	841	GTAACCTCAGTGGCCCTCAAGCCAAAGCTCTCTTCAGGATTAAGACGTAATTTGCCATCAAT	900
Qy	301	ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn	320
Db	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATTAAGACGTAATTTGCCATCAAT	960
Qy	321	GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp	340
Db	961	CAGGACCCCTTGGGAAGCAAGGGTACCAGCTTTAGACAGGGAGACAACCTTTGAAGTGTGG	1020
Qy	341	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly	360
Db	1021	GAACGACCTCTCTCAGGCTTAGCGCTTAGCGCTGTAGCTATGATAAACCAGCAGAGATTGGT	1080
Qy	361	GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro	380
Db	1081	GGACCTCGCTCTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT	1140
Qy	381	AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr	400
Db	1141	GCCTGCTTCATCACAGCTCTCTCCTGTGAAGAGAGACTAGGGTTCATGAATGGACT	1200
Qy	401	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr	420
Db	1201	TCAGGTTAAGAGTCACATTAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA	1260
Qy	421	Met 421	
Db	1261	ATG 1263	
RESULT 9			
AR654960			
LOCUS			
DEFINITION			
SEQUENCE 3 from patent US 6890748.			
ACCESSION			
VERSION			
AR654960.1 GI:67586497			
KEYWORDS			
SOURCE			
Unknown.			
ORGANISM			
Unclassified.			
REFERENCE			
1 (bases 1 to 1290)			
AUTHORS			
Garger S.J., Turpen, T.H. and Kumagai, M.H.			
TITLE			
Production of lysosomal enzymes in plants by transient expression			
JOURNAL			
Patent: US 6890748-A 3 10-MAY-2005;			
Large Scale Biology Corporation; Vacaville, CA			
FEATURES			
Location/Qualifiers			
1..1290			
/organism="unknown"			

ORIGIN		/mol_type="genomic DNA"	
Alignment Scores:			
Pred. No.:	1.37e-207	Length:	1290
Score:	2314.00	Matches:	421
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0
US-10-602-219-12 (1-421) x AR654960 (1-1290)			
Qy	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC	60
Qy	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
Db	61	CTCGTTTCTCGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT	120
Qy	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
Db	121	ACCATGGGCTGGCTGCATCTGGGAGCGCTTCAITGTGCAACCTTGACTGCCAGGAAGGCCA	180
Qy	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
Db	181	GATTCCTGCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240
Qy	81	TrpLysAspAlaGlyTyrGluTyrLeuGlyCysIleAspAspCysTrpMetAlaProGlnArg	100
Db	241	TGGAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCCAAGA	300
Qy	101	AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
Db	301	GATTTCAGAAAGCAGACTTCAGCGAGACCCCTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA	360
Qy	121	AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys	140
Db	361	GCTAATTAATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGCAGATGTTTGGAAATAAA	420
Qy	141	ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla	160
Db	421	ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCATGATGCCAGACCTTTGCT	480
Qy	161	AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu	180
Db	481	GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGTTTGGAAAATTGG	540
Qy	181	AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr	200
Db	541	GCAGATGGTTATAGCACATGTCCTTGGCCCTGAATAGACTGGCAGGAAGCAATTTGTGTAC	600
Qy	201	SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg	220
Db	601	TCCTGTGAGTGGCCCTCTTATATGTGGCCCTTCAAAAGCCCAATATATACAGAAATCCGA	660
Qy	221	GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys	240
Db	661	CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTGGAAAAGTATAAAG	720
Qy	241	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly	260
Db	721	AGTATCTTGGACTGGACATCTTTTAACCCAGGAGAGAAATTTGTGTGCTGGACCAAGG	780
Qy	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln	280
Db	781	GGTTGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCCTCAGCTGGAAATCAGCAA	840
Qy	281	ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu	300
Db	841	GTAACCTCAGATGGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTTCATGTCTAATGACCTC	900


```

Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGAGCATTAATGGCCATCAAT 960

Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGGACAGCAGGATACAGCTTAGACAGGAGACACTTTGAAGTGTGG 1020

Qy 341 GlnArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCAGCAGGAGATTGGT 1080

Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTCTGCTTCCCTGGGTAAGGAGTGGCTGTAAATCCT 1140

Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCTCTGTGAAAAGAGAGCTAGGGTTTCTATGAATGGACT 1200

Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnGluAsnThr 420
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGTAGAAAACACA 1260

Qy 421 Met 421
Db 1261 ATG 1263

RESULT 10
LOCUS BT006864 1290 bp mRNA linear PRI 13-MAY-2003
DEFINITION Homo sapiens galactosidase, alpha mRNA, complete cds.
ACCESSION BT006864
VERSION BT006864.1 GI:30582566
KEYWORDS FLI CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1290)
Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
Cloning of human full-length cDNAs in BD Creator(TM) System Donor
vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1290)
Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
COMMENT This cDNA clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each CDS has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The CDS has been directionally cloned using BD In-Fusion(TM)
cloning system between the SalI and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after SalI site
and before 'ATG' to provide Kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES
source
1..1290
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GH00112X1.0"
/clone_lib="BD Creator(TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha T1 resistant"

```

```

/note="Vector: pDNR-Dual"
1..1290
/codon_start=1
/product="galactosidase, alpha"
/protein_id="AAP35510.1"
/db_xref="GI:30582567"
/translation="MQLRNPRLHGLCALRLFLALVSWDIPGARALDNLGLARTPTMGW
LWRFNCLDQESPDCISEKLFMEWALVSEGWKADAGTEYLICIDCWNPQRDS
EGRLQADPQRFPHGIRQLANYVHSGKLGIYADVGNKTCAGFPGSFYDIDAQTF
DWGVLLKFDGCGYCDLSLADYKRMSLALNRGTSIVSCSEFLYVMPFQKPNYTE
IROYCNWRNFADIDDSWKSISLDWTSFNQERIVDVAGPGWNDPMLVTGNFGLS
WNOQVOTOMALWAINAAPLEFMSNDRIHISPOAKALLQDKDVIAINQDPLGKGGYQLRQG
DNFVWRPDLGLAWAVAMINRQISGRPSYTIASVSLGKGVACNPACFITQLLPVKR
KLGFYEWTSRLRSHINPTGVLLQLENTMQMSLKDLL"
ORIGIN
Alignment Scores:
Pred. No.: 1,37e-207 Length: 1290
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 8
US-10-602-219-12 (1-421) x BT006864 (1-1290)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTCGTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGGTTTCTGGGACATCCCTGGGCTAGACACTGGCAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGGCTGGAGCGCTTCATGTGCAACCTTGACTGCAGGAAGAGCCA 180
Qy 61 AspSerCysIleSerGlyLeuPheMetGluMetAlaGluLeuMetValSerGly 80
Db 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGACAGGTTATGAGTACCTCTGTCATTGATGACTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGlyValArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGGAGGACAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGTCA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGAGGACTGAAGCTAGGATTTATGCAGATGTTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTCCCTCGGGAGTTTGGTACTACAGCATTTGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTATAGCACATGTCCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeArg 220
Db 601 TCCTGTGATGGCTCTTTATATATGTGGCCCTTTTCAAAGAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTCAATCACTTCGCGAAATTTTGTGACATTTGATGATTCCTCGAAAGATTAAG 720

```

QY	241	SerIleLeuAspTrpThrSerPhehGlnGluAArgIleValaspValaLaGlyProGly	260
Db	721	AGTATCTTGGACTGGACATCTTTTAACCCAGGAGAAATTTGTATGTTGCTGGACAGGG	780
QY	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln	280
Db	781	GTTTGGATGACCCAGATATGTAGTATGGCACTTTGGCTCAGCTGGATCAGCNA	840
QY	281	ValThrGlnMetAlaLeuTrpAlaIleMetAlaLaProLeuPheMetSerAsnAspLeu	300
Db	841	GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC	900
QY	301	ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn	320
Db	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT	960
QY	321	GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAsnPheGluValTrp	340
Db	961	CAGGACCCCTTGGCCAGCAGGGTACCAGCTTAGACAGGAGACAACTTTGAGTGTGG	1020
QY	341	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly	360
Db	1021	GAACGACCTCTCTCAGCTTAGCTGGGCTGTAGCTATGATAAACCCGGCAGGAGATTGCT	1080
QY	361	GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro	380
Db	1081	GGACCTGGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCCCTGTAATCCT	1140
QY	381	AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr	400
Db	1141	GCTGCTTCATCACAGCTCTCCTCTGTGAAGAAGTAGGTTCTATGAATGGACT	1200
QY	401	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr	420
Db	1201	TCAAGGTTAAGAAGTCACATAATCCACAGGCACCTGTTTCTTCTCAGCTAGAAAATACA	1260
QY	421	Met 421	
Db	1261	ATG 1263	
RESULT 11			
AY889631			
LOCUS			
DEFINITION	AY889631.1	1290 bp mRNA linear SYN 29-MAR-2005	
ACCESSION		Synthetic construct Homo sapiens clone FLH025769.01X galactosidase	
VERSION		alpha (GLA) mRNA, complete cds.	
KEYWORDS		AY889631	
SOURCE		Human ORF Project.	
ORGANISM		synthetic construct	
REFERENCE		other sequences; artificial sequences.	
AUTHORS		1 (bases 1 to 1290)	
		Shen, B., Halleck, A., Jenson, D., Moreira, D., Raphael, J., Kelley, F.,	
		Williamson, J. and Labaer, J.	
TITLE		Cloning of human full-length CDS in Creator (TM) recombinational	
JOURNAL		vector system	
REFERENCE		Unpublished	
AUTHORS		2 (bases 1 to 1290)	
		Hines, L., Rolfe, A., Jenson, D., Moreira, D., Raphael, J., Kelley, F.,	
		Shen, B., Halleck, A., Koundinya, M., Hu, Y., Zuo, D., Taycher, E.,	
		Williamson, J. and Labaer, J.	
TITLE		Direct Submission	
JOURNAL		Submitted (04-JAN-2005) Biological Chemistry and Molecular	
COMMENT		Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,	
		Cambridge, MA 02141, USA	
		This CDS clone is a part of a collection of human full-length	
		expression clones generated by Harvard Institute of Proteomics.	
		This ORF clone has been cloned with normalized stop-codon. The CDS	
		has been directionally cloned using BD In-Fusion(TM) cloning system	
		between the SalI and HindIII sites of the pONR-Dual vector.	
		Additional sequences in the clone: 'ACC' after SalI site and	

before 'ATG' to provide Kozak consensus sequence. Each clone is		clonally isolated and full-length sequence-verified.	
FEATURES		Location/Qualifiers	
source		1..1290	
		/organism="synthetic construct"	
		/mol_type="mRNA"	
		/db_xref="taxon:32630"	
		/clone="FLH025769.01X"	
		/lab_host="Escherichia coli DH5alpha Tl resistant"	
		/note="derived from MGC template"	
gene		1..1290	
		/gene="GLA"	
CDS		1..1290	
		/gene="GLA"	
		/codon_start=1	
		/transl_table=11	
		/product="galactosidase alpha"	
		/protein_id="AA32422.1"	
		/db_xref="GI:60655717"	
		/translation="MQLNPRLHLCALALRFLALVSWDIPCARALDNLARTPTMGW	
		LHWRFCNLDQCQEPDSCISEKLFMEAEMLVSEGWKADAGEYLICIDDCWAPQD	
		EGRLQADQRFPHGIRQLANYVHSGKLKGIYADVGNKTCAGFPGSFYDIDAQTF	
		DWGVDLLKFDGCDLSLENLADYVHMSLALNRTGRSIVYSCWLYMPPQKPNY	
		IRQYCNWRNFADIDDSWKSISILDWTSFNQERIVDVAGPGWMDPLMVLITGNF	
		GLS WNOQYTMALWAIWAAPLFMNSDLRHSIQAKALQDKDVIATINQDPLGKQGY	
		LRQG DNFVWERPLSLAWAVAMINROEIGGPRSYTIAVASLGKGVACNPACFITQL	
		LPVKR KLGFYEWTSRLRSHINPTGTVLLQLLENTMQMSLKDLL"	
ORIGIN			
Alignment Scores:			
Pred. No.:		1.37e-207	
Score:		2314.00	
Length:		1290	
Matches:		421	
Percent Similarity:		100.00%	
Conservative:		0	
Best Local Similarity:		100.00%	
Mismatches:		0	
Query Match:		100.00%	
Indels:		0	
DB:		11	
Gaps:		0	
US-10-602-219-12 (1-421) x AY889631 (1-1290)			
QY	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
Db	1	ATGCAGCTCAGGAACCCAGAACTACATCTGGGCTGCGGCTTTCGCTTCTCTGGGC	60
QY	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
Db	61	CTCGTTTCTTGGGACATCCTCGGGCTAGACACCTGGCAATGGATTGGCAAGGACCT	120
QY	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
Db	121	ACCATGGGCTGGCTGCATCGGAGCGCTTCATGTCAACCTTGACTGCCAGGAAGCCA	180
QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
Db	181	GATTCCTGCATCAGTCAGAGAAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGA	240
QY	81	TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg	100
Db	241	TGGAAGGATGCAAGTTATGATGACCTCTCATGTGATGACTGTGTGATGCTCCCCA	300
QY	101	AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
Db	301	GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCC	360
QY	121	AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys	140
Db	361	GCTAATATTATGTTTACACAGCAAGGACCTGAAGCTAGGGAATTTATGCAGATGTT	420
QY	141	ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla	160
Db	421	ACCTGCGCAGGCTTCCCTGGAGTTTGGATTAAGTACTAGACATTTGATGCCAGAC	480
QY	161	AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu	180

Db 481 CACTGGGAGTAGATCTCTAAATTTGATGGTTGTTACTGTGACAGTTTCGAAAATTG 540
 Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAenArgThrGlyArgSerIleValTyr 200
 Db 541 GCAGATGTTATAGACATGCTCTTGGCCCTGAATAGACTGGCAGAAAGCATTTGTGTAC 600
 Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 Db 601 TCCTGTGAGTGGCCCTCTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
 Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTGCATCACTGGCGAAATTTTCTGACATGTATGATTTCTGGAAAGATATAAG 720
 Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTATGTTGCTGGACAGGG 780
 Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
 Db 781 GGTGGAAATGACCCAGATATGTATGTATGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
 Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
 Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGAGTAAATGGCATCAAT 960
 Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGGCAAGAGGTACCAAGCTTAGACAGGGAGAGCAACTTTGAAGTGTGG 1020
 Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATGAATAAACCGCAGAGATTGGT 1080
 Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
 Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAAGTGGTTCATGAATGGACT 1200
 Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 Db 1201 TCNAAGTTAAGAAGTCAATAAATCCACAGGCACACTGTTTTCCTTCAGCTAGAAATACA 1260
 Qy 421 Met 421
 Db 1261 ATG 1263

RESULT 12
 AY889632
 LOCUS
 DEFINITION Synthetic construct Homo sapiens clone FLH025770.01x galactosidase alpha (GLA) mRNA, complete cds.
 ACCESSION AY889632
 VERSION AY889632.1 GI:60655718
 KEYWORDS Human ORF project.
 SOURCE synthetic construct
 ORGANISM synthetic construct
 other sequences; artificial sequences.
 REFERENCE 1 (bases 1 to 1290)
 AUTHORS Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
 Williamson,J. and LaBaer,J.
 Cloning of human full-length CDS in Creator (TM) recombinational vector system
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1290)

AUTHORS Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
 Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
 Williamson,J. and LaBaer,J.
 TITLE Direct Submission
 JOURNAL Submitted (04-JAN-2005) Biological Chemistry and Molecular
 Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
 Cambridge, MA 02141, USA
 COMMENT This CDS clone is a part of a collection of human full-length
 expression clones generated by Harvard Institute of Proteomics.
 This ORF clone has been cloned with normalized stop-codon. The CDS
 has been directionally cloned using BD In-Fusion(TM) cloning system
 between the SalI and HindIII sites of the pDNR-Dual vector.
 Additional sequences in the clone: 'ACC' after SalI site and
 before 'ATG' to provide Kozak consensus sequence. Each clone is
 clonally isolated and full-length sequence-verified.
 FEATURES
 Location/Qualifiers
 1..1290
 /organism="synthetic construct"
 /mol_type="mRNA"
 /db_xref="taxon:32630"
 /clones="FLH025770.01x"
 /lab_host="Escherichia coli DH5alpha T1 resistant"
 /note="derived from MGC template"
 gene
 1..1290
 /gene="GLA"
 CDS
 1..1290
 /codon_start=1
 /transl_table=11
 /product="galactosidase alpha"
 /protein_id="AA32423.1"
 /translation="MQLRNPDLHLGSCALALFLALVSWDIPGARALDNLARTPTMGW
 LHWRFNCLDCQEPDSCISEKLFMEAMELMWSEGWKADGVEYLICIDCWAPQDTS
 EGRLOADPQRPFGHIGLRYLVVHSGKLGIYADVGNKTCAGFPGSFGYDIDAOTFA
 DWGYDLLKFGCYCDLSENLADGYKMSLANRTGRSVYSCWPLNWPQKPNVTE
 IRQCNWRFPADIDDSKWSIKSLDWTSTFNQERIVDVAGGGWNPDLMLVGNFGLS
 WNQQTQMAIWAIMAAFLFMSNDLHSPQAKLLQDKVDVAINDDPGKGGYQURQG
 DNFEVWRPLSGLAWAMINRQBIGRPSTIAVASLKGKGVACNPACFITQLLPVKR
 KLGFYEWTSRLRSHINPTGTVLLQLENTMQMSLKDLL"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.37e-207 Length: 1290
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 11
 US-10-602-219-12 (1-421) x AY889632 (1-1290)
 Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGAGAACCCAGAACTACATCTGGCTGCGCGCTTGCCTTCCTCTGCGC 60
 Qy 21 LeuValSerTrpAspIleProGlyValAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGACACTGGACAATGGATTGGCAAGGACGCT 120
 Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCAGAGAGCCA 180
 Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCTCTCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGTCTCTCAGAAGGC 240
 Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAAAGGATGAGGTTATGATGACTCTCTGATGATGATGATGATGATGATGATGATG 300
 Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120

Db 301 GATTTCAGAAAGGAGAGCTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCCGCAGCTA 360
Qy 121 AlaAenTyValHiserlysglyLeuylsleuGlylleTyAlaaspValglyAsnlys 140
Db 361 GCTAATATATGTCACAGCAAGAGCTGAAGCTAGGGAATTAATGAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyTyTyAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGAGCTTCCTGGGAGTTTTGGATACTACGATTCATGATGCCAGACCTTTGCT 480
Qy 161 AspTTPGlyValAspLeuLeuylsPheAspGlyCysTyTyCysAspSerleuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGATTTGGAAATTTG 540
Qy 181 AlaAspGlyTyTyTyHisMetSerleuAlaLeuAenArgThrGlyArgSerIleValTy 200
Db 541 GCAGATGGTTATPAGCACATGCTTGGCCCTGAAATAGGACTGGCAGAAAGCAATGTGTAC 600
Qy 201 SerCysGluTTPProLeuTyMetTTPProPheGlnLysProAsnTyThrGluIleArg 220
Db 601 TCTGTGAGTGCCCTCTTTATATGTGGCCCTTCAAAAGCCCAATATACAGAAATCCGA 660
Qy 221 GlnTyTyCysAsnHisTTPArgAsnPheAlaAspIleAspAspSerTTPlysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTCATGATTCCTGGAAAGATATAAG 720
Qy 241 SerIleLeuAspTTPThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAATTTGTGATGTTGCTGGACACGG 780
Qy 261 GlyTTPAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTTPAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGATCAGCA 840
Qy 281 ValThrGlnMetAlaLeuTTPAlaIleMetAlaIaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 GCACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATGGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyTyGlnLeuArgGlnGlyAspAsnPheGluValTTP 340
Db 961 CAGGACCCCTTGGGCAGCAGGATACCGCTTAGACGGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTTPAlaValAlaMetIleAsnArgGlnGluLeGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCCGGCAGGAGATTGT 1080
Qy 361 GlyProArgSerTyThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTATACCATCCAGTGTGCTTCCCTGGGTAAAGGAGTGGCCTGTAACTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyTyGluTTPThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCTGTGAAGAAGAGCTAGGTTTCTATGAAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTTGCTTCAGCTAGAAAATACA 1260
Qy 421 Met 421
Db 1261 ATG 1263

RESULT 13
AY892090
LOCUS
DEFINITION
alpha (GLA) mRNA, partial cds.
ACCESSION
AY892090

AY892090.1 GI:60652624
Human ORF project.
SYNTHETIC CONSTRUCT
SYNTHETIC CONSTRUCT
OTHER SEQUENCES; ARTIFICIAL SEQUENCES.
1 (bases 1 to 1290)
Hines,B., Rolfs,A., Jepsen,D., Moreira,D., Raphael,J., Kelley,F.,
Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
Williamson,J. and LaBaer,J.
Cloning of human full-length CDS in Creator (TM) recombinational
vector system
Unpublished
2 (bases 1 to 1290)
Hines,B., Rolfs,A., Jepsen,D., Moreira,D., Raphael,J., Kelley,F.,
Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
Williamson,J. and LaBaer,J.
Direct Submission
Submitted (05-JAN-2005) Biological Chemistry and Molecular
Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
Cambridge, MA 02141, USA
This CDS clone is a part of a collection of human full-length
expression clones generated by Harvard Institute of Proteomics.
This ORF clone has been cloned without stop-codon (to allow fusion
with C-terminal tag). The CDS has been directionally cloned using
BD In-Fusion(TM) cloning system between the SalI and HindIII sites
of the pDNR-Dual vector. Additional sequences in the clone: 'ACC'
after SalI site and before 'ATG' to provide Kozak consensus
sequence; 'GG' after last codon and before HindIII site to maintain
reading frame. Each clone is clonally isolated and full-length
sequence-verified.
Location/Qualifiers
1. 1290
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clones="FLH025766.01L"
/lab_host="Escherichia coli DH5alpha T1 resistant"
/note="derived from MGC template"
1. >1290
/gene="GLA"
1. >1290
/gene="GLA"
/codon_start=1
/transl_table=11
/product="galactosidase alpha"
/protein_id="AA29007.1"
/db_xref="GI:60652625"
/translation="MQLRNPHELHGCALRFLALVSWDIPGARLDNGLARTPTMGW
LHWRFMCNLDQEPDSCISEKLFMEAMELMSEGWKADGYEYLICDDCWAPQDS
EGRLQADPQRPHGIROLANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTF
DWGVDLLKFDGDCYDSLNLADGVKHMSLALNRTGRSIVYSCFWPLVWPKPNYTE
IRQCNHWRNFADIDDSKSIKSLDTSFNQERIVDVAGGWNDDPMLVIGNGLS
NNQVQTQALWAIWAAPLUFMSNDLRHISFQAKLLQDKDVIADINQDPLGKQGYLRQR
DNFEWVERPUSGLAWAMINRQETIGRPSRTIAVASLGKGVACNPACFTTQLLPVKR
KLGFVETSRLSRSHINPTGTVLLQLENTMQMSLKDLLL"
Alignment Scores:
Pred. No.: 1.37e-207 Length: 1290
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Db: 11 Gaps: 0
US-10-602-219-12 (1-421) x AY892090 (1-1290)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTGGCTTCGCTTCTCGGCC 60
Qy 21 LeuValSerTTPAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40

Db 61 CTCGTTCTCTGGGACATCCCTGGGGCTAGAGCAGCTGGACAATGATGGCAAGGAGCGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCTGCATCAGTGAGAAGCTCTTCATGTGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGAGGTTATGAGTACCTCTGCATGTGATGACTGTGTGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleAArgGlnLeu 120
Db 301 GATTGAGAAGCAGCAGCTTCAGGAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaIleTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGGACTGAAGCTAGGAGTATTATGCAGATGTTTGGAAATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACGACATGTATGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTG 600
Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTCGGCCCTTCAAAAGCCCAATATATACAGAAATCGA 660
Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCATCAGTGGGAAATTTGCTGACATGTGATGATTTCTGGAAAGTATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTAAACAGGAGAGAAATTTGATGTTGCTGGACACGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGTATGGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTTACAGGATAAGGACGTAATTTGCCATCA 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTAGACAGGGAGAGCAACTTTTGAAGTGTG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAAACCGGACGAGATTTGGT 1080
Qy 361 GlyProArgSerTrpIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCCTGCTTTCATCACAGCTCTCTCCCTGTGAAAGGAGGCTAGGCTTCTATGAATGGACT 1200

Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTAAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTTCAGTAGAAAATACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
RESULT 14
LOCUS BT007835 1290 bp mRNA linear SYN 13-MAY-2003
DEFINITION Synthetic construct Homo sapiens galactosidase, alpha mRNA, partial cds.
ACCESSION BT007835
VERSION BT007835.1 GI:30584508
KEYWORDS FLI CDNA.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Kalline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y., Phelan,M. and Farmer,A.
TITLE Cloning of human full-length CDSs in BD Creator (TM) System Donor vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1290)
AUTHORS Kalline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y., Phelan,M. and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame.
Clone distribution: <http://bioinfo.clontech.com/orfclones>.
FEATURES
Location/Qualifiers
1..1290
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="GH00112L1.0"
/clone_lib="BD Creator (TM) CDS Library derived from MGC collection"
/lab_host="DH5alpha T1 resistant"
/note="vector: pDNR-Dual"
1..>1290
/note="Mutations: 1289:Stop->Leu"
/codon_start=1
/transl_table=1
/product="Homo sapiens galactosidase, alpha"
/protein_id="AAP36507.1"
/db_xref="GI:30584509"
/translation="MQLRNPELHIGCALALRFLALVSWDIFGARALDNLGARTPTMGW LHWRFNCLDCEPDSICSEKLFMEAMLMVSEGWKADGYEYLICIDCWMAPOKDS EGRLOADPQPFPHGIROLANYVHSGKLGIYADVGNKTCAGFPGSFGYDIDACTEA DWGVLLKFGGCTCSDLENADGYSKMSLALNRGRSIVYSCENPLNWPQKPNYTE IROYCNHRNFADIDDSWKSISKILDTSFNQRIVDVAGGSNDPMLVLGNFGLS WNOQTQMLAWMAAPLFMSNDLRHSIPQAKLLQDKDVTAINDPGLKQGYLQROG DNFVWRPLSLGAWAMINRQISGIPRSYTIIVASLKGKGVACNPACFITQLLPVKR KLGFEWTSRLSRHINPTGTVLLQLENTMQMSLKDLLL"
ORIGIN
Alignment Scores:
Pred. No.: 1.37e-207 Length: 1290

Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0

US-10-602-219-12 (1-421) x BT007835 (1-1296)

QY 1 MetGlnLeuArgAspProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTCGCTTCCTGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCATTGGACATGGATTGGCAAGAGCGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAGGATGAGGTATGAGTACTCTGCAATTGATGACTGTTGGATGGCTCCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTGAGAAGGCAGACTTCAGGCAGACCTCAGCGCTTCTCTCATGGGATTCGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerIleGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATTATGTTTCAGCAGCAAGAGCTGAAGCTAGGATTTATGCAAGATGTTGGAATAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGGTTCCTGGGAGTTTGGATACTACGACATGTGATGCCAGACCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTTGTACTGTGACAGATTTGGAATAATTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 CGAGATGGTTATAGCACATGCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTCAAGGCCCAATTTATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTGTGTCATTTGATGATTCCTGGAAAGATATAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValaGlyProGly 260
DB 721 AGTATCTGGACTGGACATCTTTTAACCCAGGAGAGAAATTTGTGATGTTGCTGGACAGGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGATGACCCAGATATGTTAGTATGGCACTTTGGCTTCAGCTGGATCAGCNA 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900

QY 301 ArgHisIleSerProGlnLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960

QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340

DB 961 CAGGACCCCTGGGCAAGAGGTTACAGCTTAGACAGGAGAGCAACTTTGAAGTGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCCGAGAGATTGGT 1080

QY 361 GlyProArgSerTyrThrIleAlaValaAlaSerLeuGlyLysGlyValaAlaCysAsnPro 380
DB 1081 GGACTCGCTTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1141 GCCTGCTTATCATCACAGCTCCTCCCTGTGAAAAGGAAGCTAGGCTTCTATGAATGGACT 1200

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1201 TCAAGGTTAAGAAGTCACATAAATCCCAAGGCACTGTTTTGCTTCAGCTAGAAAATACA 1260

QY 421 Met 421
DB 1261 ATG 1263

RESULT 15
AR653695 LOCUS AR653695 1296 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 9 from patent US 6887696.
ACCESSION AR653695
VERSION AR653695.1 GI:67584280
KEYWORDS SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1296)
AUTHORS Garg,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 9 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source 1..1296
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 1,38e-207 Length: 1296
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-219-12 (1-421) x AR653695 (1-1296)

QY 1 MetGlnLeuArgAspProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTCGCTTCCTGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCATTGGCAATGGATTGGCAAGAGCGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAGGATGAGGTATGAGTACTCTGCAATTGATGACTGTTGGATGGCTCCCCAAGA 300

Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db |||||
301 GATTGAGAAGGCAGACATTCAGGCAGACCTCAGGCTTTCCTCATGGGATTGCGCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db |||||
361 GCTAATTATGTTTACAGCAAGGACTGAAGCTAGGGATTTATGAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db |||||
421 ACCTGGCAGAGCTTCCCTGGGAGTTTGGATACACGACATTTGATGCCACACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db |||||
481 GACTGGGAGTAGATCTGCTTAAATTTTGATGGTTGTACTGTGACAGTTTGGAAAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db |||||
541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db |||||
601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTCAAAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db |||||
661 CAGTACTGCAATCACTGGCGAAATTTTGCTGACATTTGATGATTCCTGGAAAAATATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db |||||
721 AGTATCTTGGACTGGACATCTTTAAACCAGGAGAGATTTGTGATTTGCTGGACCCAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db |||||
781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db |||||
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTCATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaIysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db |||||
901 CGACACATCACCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAAATGGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db |||||
961 CAGGACCCCTTGGGCAAGCAGGGTACCGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db |||||
1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCCGCGAGGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db |||||
1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db |||||
1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAGAAGAGCTAGGGTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db |||||
1201 TCAAGTTTAAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTTCAGCTAGAAAAACACA 1260
Qy 421 Met 421
Db |||
1261 ATG 1263

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 24, 2005, 08:58:24 ; Search time 44 Seconds
(without alignments)
791.056 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MQLNPELHGCALALRFLA.....RLRSHINPTGTVLLQLENTM 421

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA*
1: /cgn2_6/ptodata/1/iaa/5-COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6-COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H-COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCUTUS-COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE-COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	421	2	US-09-993-059-12
2	2314	100.0	421	2	US-10-103-327-12
3	2314	100.0	427	2	US-09-993-059-14
4	2314	100.0	427	2	US-10-103-327-14
5	2314	100.0	429	1	US-07-602-824A-2
6	2314	100.0	429	1	US-07-602-608-3
7	2314	100.0	429	1	US-07-983-451-2
8	2314	100.0	429	1	US-08-261-578-3
9	2314	100.0	429	1	US-08-261-577-7
10	2314	100.0	429	1	US-08-261-577-9
11	2314	100.0	431	2	US-09-993-059-10
12	2314	100.0	431	2	US-10-103-327-10
13	2314	100.0	435	2	US-09-993-059-6
14	2314	100.0	435	2	US-10-103-327-6
15	2310	99.8	429	2	US-09-070-356-4
16	2309	99.8	424	2	US-09-993-059-8
17	2309	99.8	424	2	US-10-103-327-8
18	2309	99.8	428	2	US-09-993-059-4
19	2309	99.8	428	2	US-10-103-327-4
20	2296	99.2	428	6	5179023-4
21	2293	99.1	417	2	US-09-993-059-16
22	2293	99.1	417	2	US-10-103-327-16
23	2293	99.1	423	2	US-09-993-059-18
24	2293	99.1	423	2	US-10-103-327-18
25	2226	96.2	409	2	US-09-993-059-22
26	2226	96.2	409	2	US-10-103-327-22
27	2213	95.6	401	2	US-09-993-059-20

28	2213	95.6	401	2	US-10-103-327-20	Sequence 20, Appl
29	2152	93.0	390	2	US-09-176-666-7	Sequence 7, Appl
30	2152	93.0	391	2	US-09-176-666-6	Sequence 6, Appl
31	2152	93.0	392	2	US-09-176-666-5	Sequence 5, Appl
32	2152	93.0	393	2	US-09-176-666-4	Sequence 4, Appl
33	2152	93.0	394	2	US-09-176-666-3	Sequence 3, Appl
34	2152	93.0	396	2	US-09-176-666-2	Sequence 2, Appl
35	2152	93.0	398	2	US-08-928-881-26	Sequence 26, Appl
36	2152	93.0	398	2	US-09-176-666-1	Sequence 1, Appl
37	2152	93.0	398	2	US-09-543-921-26	Sequence 26, Appl
38	2152	93.0	398	2	US-09-266-014-4	Sequence 4, Appl
39	2152	93.0	398	2	US-09-491-759-26	Sequence 26, Appl
40	2152	93.0	398	2	US-10-360-101-202	Sequence 202, App
41	2147	92.8	389	2	US-09-176-666-8	Sequence 8, Appl
42	2142	92.6	388	2	US-09-176-666-9	Sequence 9, Appl
43	2136	92.3	387	2	US-09-176-666-10	Sequence 10, Appl
44	2131	92.1	386	2	US-09-176-666-11	Sequence 11, Appl
45	2110	91.2	381	2	US-09-176-666-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-12
; Sequence 12, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-12

Query Match	100.0%	Score 2314;	DB 2;	Length 421;
Best Local Similarity	100.0%	Pred. No. 1.4e-234;		
Matches 421;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MQLNPELHGCALALRFLALVSWDIPGARALDNLGARTPTMGMLHWRPFCNLDQSEP	60	
Db	1	MQLNPELHGCALALRFLALVSWDIPGARALDNLGARTPTMGMLHWRPFCNLDQSEP	60	
Qy	61	DSCTSEKLFMEAMLSYSEGWKADGAYEYLCTDDCWAPORSEGLQADPQPFPHGIRQL	120	
Db	61	DSCTSEKLFMEAMLSYSEGWKADGAYEYLCTDDCWAPORSEGLQADPQPFPHGIRQL	120	
Qy	121	ANYVHSGKLGIYADVGNKTCAGPSPGFGYDIDAQTFADGWVDLLKFDGCGYCDSENL	180	
Db	121	ANYVHSGKLGIYADVGNKTCAGPSPGFGYDIDAQTFADGWVDLLKFDGCGYCDSENL	180	
Qy	181	ADGKHSMLALNRTGRSIVYSCWPLMYMPFKPNYTEIRQYCNHRNFADIDDSWKSIIK	240	
Db	181	ADGKHSMLALNRTGRSIVYSCWPLMYMPFKPNYTEIRQYCNHRNFADIDDSWKSIIK	240	
Qy	241	SILDWTSFNQRIYDVAGPGGNDPDMLVINGFGLSNQVQVTOMALWAIMAAPLPMNDL	300	
Db	241	SILDWTSFNQRIYDVAGPGGNDPDMLVINGFGLSNQVQVTOMALWAIMAAPLPMNDL	300	
Qy	301	RHISPOKALLQDKDVIAINODPLGKQYQLRGDNPEVWERPLSGLAWAVAMINRQIG	360	
Db	301	RHISPOKALLQDKDVIAINODPLGKQYQLRGDNPEVWERPLSGLAWAVAMINRQIG	360	
Qy	361	GPRSYTIAVSLGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT	420	

||||| 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Db
QY 421 M 421
|
Db 421 M 421

RESULT 2

US-10-103-327-12
; Sequence 12, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-12

Query Match 100.0%; Score 2314; DB 2; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.4e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFALVSWDIPGARALDNLGLARTPTMGWLHWFRCMNCDCQSEP 60
Db 1 MQLRNPGLHGCALALRFALVSWDIPGARALDNLGLARTPTMGWLHWFRCMNCDCQSEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPORDSEGRLOADPQPFPHGIROL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPORDSEGRLOADPQPFPHGIROL 120
QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGYGYDDAQTADWGVLLKFDGCGYCDLSLENT 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGYGYDDAQTADWGVLLKFDGCGYCDLSLENT 180
QY 181 ADGYKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROQCNHWRNFADIDDSWKSIX 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROQCNHWRNFADIDDSWKSIX 240
QY 241 SILDWTSFNQERIVDVAGPGWNPDMVLVGNFGLSNQOVTQMALWAIMAAPLFWMSNDL 300
Db 241 SILDWTSFNQERIVDVAGPGWNPDMVLVGNFGLSNQOVTQMALWAIMAAPLFWMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
|
Db 421 M 421

RESULT 3

US-09-993-059-14
; Sequence 14, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-14

Query Match 100.0%; Score 2314; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFALVSWDIPGARALDNLGLARTPTMGWLHWFRCMNCDCQSEP 60
Db 1 MQLRNPGLHGCALALRFALVSWDIPGARALDNLGLARTPTMGWLHWFRCMNCDCQSEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPORDSEGRLOADPQPFPHGIROL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPORDSEGRLOADPQPFPHGIROL 120
QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGYGYDDAQTADWGVLLKFDGCGYCDLSLENT 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGYGYDDAQTADWGVLLKFDGCGYCDLSLENT 180
QY 181 ADGYKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROQCNHWRNFADIDDSWKSIX 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROQCNHWRNFADIDDSWKSIX 240
QY 241 SILDWTSFNQERIVDVAGPGWNPDMVLVGNFGLSNQOVTQMALWAIMAAPLFWMSNDL 300
Db 241 SILDWTSFNQERIVDVAGPGWNPDMVLVGNFGLSNQOVTQMALWAIMAAPLFWMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
|
Db 421 M 421

RESULT 4

US-10-103-327-14
; Sequence 14, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT

; FILING DATE: 24-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-07-602-608-3

Query Match 100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQDSEGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQDSEGRLOADPQRPFGIRQL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTAFADGWVLLKFDGICYDLSNLL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTAFADGWVLLKFDGICYDLSNLL 180
QY 181 ADGYKHSALNLRGTSIVYSCWPLVMPFQKPNYTEIROVCHWRNFPADIDDSWSKSIK 240
DB 181 ADGYKHSALNLRGTSIVYSCWPLVMPFQKPNYTEIROVCHWRNFPADIDDSWSKSIK 240
QY 241 SILDWTFSNQERIVDVAGPGGWNDDMLVIGNFGLSWNQVQTOMALWAIMAAPLFMSNDL 300
DB 241 SILDWTFSNQERIVDVAGPGGWNDDMLVIGNFGLSWNQVQTOMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQIEIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQIEIG 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGYEWTSLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGYEWTSLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421

RESULT 7

US-07-983-451-2
; Sequence 2, Application US/07983451
; Patent No. 5401650
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/983,451
; FILING DATE: 30-NOV-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 07/983,451
; REFERENCE/DOCKET NUMBER: 6923-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-983-451-2

Query Match 100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQDSEGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQDSEGRLOADPQRPFGIRQL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTAFADGWVLLKFDGICYDLSNLL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTAFADGWVLLKFDGICYDLSNLL 180
QY 181 ADGYKHSALNLRGTSIVYSCWPLVMPFQKPNYTEIROVCHWRNFPADIDDSWSKSIK 240
DB 181 ADGYKHSALNLRGTSIVYSCWPLVMPFQKPNYTEIROVCHWRNFPADIDDSWSKSIK 240
QY 241 SILDWTFSNQERIVDVAGPGGWNDDMLVIGNFGLSWNQVQTOMALWAIMAAPLFMSNDL 300
DB 241 SILDWTFSNQERIVDVAGPGGWNDDMLVIGNFGLSWNQVQTOMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQIEIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQIEIG 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGYEWTSLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGYEWTSLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421

RESULT 8

US-08-261-578-3
; Sequence 3, Application US/08261578
; Patent No. 5491075
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; APPLICANT: Wang, Anne M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY

TITLE OF INVENTION: ACTIVE ALPHA-N-ACETYLGLACTOSAMINIDASE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,578
FILING DATE: 17-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/602,608
FILING DATE: 24-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6923-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-261-578-3

Query Match 100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGMLHWRFMCNLDQCSEP 60
DB 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGMLHWRFMCNLDQCSEP 60

QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGLQADPQRFPHGIQRL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGLQADPQRFPHGIQRL 120

QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYVDIDAQTFADGWVLLKFDGCYCDLENL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYVDIDAQTFADGWVLLKFDGCYCDLENL 180

QY 181 ADGKXMSLALNRTGRSIVYCEWPLYWMPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGKXMSLALNRTGRSIVYCEWPLYWMPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240

QY 241 SILDWTSENRERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFWNDL 300
DB 241 SILDWTSENRERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFWNDL 300

QY 301 RHISFOAKALLQDKDVIATINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQIG 360
DB 301 RHISFOAKALLQDKDVIATINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQIG 360

QY 361 GPRSTYIIASVAGKGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
DB 361 GPRSTYIIASVAGKGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420

QY 421 M 421
DB 421 M 421

RESULT 9
US-08-261-577-7
Sequence 7, Application US/08261577
Patent No. 5580757
GENERAL INFORMATION:
APPLICANT: Desnick, Robert J.
APPLICANT: Bishop, David F.
APPLICANT: Ioannou, Yiannis A.
TITLE OF INVENTION: Cloning and Expression of Biologically
TITLE OF INVENTION: Active alpha-Galactosidase A
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,577
FILING DATE: 17-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6923-042
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 429 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-261-577-7

Query Match 100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGMLHWRFMCNLDQCSEP 60
DB 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGMLHWRFMCNLDQCSEP 60

QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGLQADPQRFPHGIQRL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGLQADPQRFPHGIQRL 120

QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYVDIDAQTFADGWVLLKFDGCYCDLENL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYVDIDAQTFADGWVLLKFDGCYCDLENL 180

QY 181 ADGKXMSLALNRTGRSIVYCEWPLYWMPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGKXMSLALNRTGRSIVYCEWPLYWMPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240

QY 241 SILDWTSENRERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFWNDL 300
DB 241 SILDWTSENRERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFWNDL 300

QY 301 RHISFOAKALLQDKDVIATINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQIG 360
DB 301 RHISFOAKALLQDKDVIATINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQIG 360

QY 361 GPRSYTIAVASLGKGVACNPACFIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Db |||||||
361 GPRSYTIAVASLGKGVACNPACFIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 10

US-08-261-577-9
; Sequence 9, Application US/08261577
; Patent No. 5580757
; GENERAL INFORMATION:
; APPLICANT: Deenick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,577
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-042
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-261-577-9

Query Match 100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEP 60
Db |||||||
1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYVLCIDDCWMAQRDSEGRQADPQRFPHGIRQL 120
Db |||||||
61 DSCISEKLFMEAEMLVSEGWKADAGEYVLCIDDCWMAQRDSEGRQADPQRFPHGIRQL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGFYGYDDAQTADWGVLDLLKFDGCGYCDSENL 180
Db |||||||
121 ANYVHSGKGLGIYADVGNKTCAGPFGSGFYGYDDAQTADWGVLDLLKFDGCGYCDSENL 180
QY 181 ADGYKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSITK 240
Db |||||||
181 ADGYKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSITK 240
QY 241 SILDWTSEFQRIYDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300

Db |||||||
241 SILDWTSEFQRIYDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQDGNFEVWERPLSGLAWAVAMINRQETG 360
Db |||||||
301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQDGNFEVWERPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVASLGKGVACNPACFIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Db |||||||
361 GPRSYTIAVASLGKGVACNPACFIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
Db 421 M 421
RESULT 11
US-09-993-059-10
; Sequence 10, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-10

Query Match 100.0%; Score 2314; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEP 60
Db |||||||
1 MOLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYVLCIDDCWMAQRDSEGRQADPQRFPHGIRQL 120
Db |||||||
61 DSCISEKLFMEAEMLVSEGWKADAGEYVLCIDDCWMAQRDSEGRQADPQRFPHGIRQL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGFYGYDDAQTADWGVLDLLKFDGCGYCDSENL 180
Db |||||||
121 ANYVHSGKGLGIYADVGNKTCAGPFGSGFYGYDDAQTADWGVLDLLKFDGCGYCDSENL 180
QY 181 ADGYKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSITK 240
Db |||||||
181 ADGYKHMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSITK 240
QY 241 SILDWTSEFQRIYDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Db |||||||
241 SILDWTSEFQRIYDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQDGNFEVWERPLSGLAWAVAMINRQETG 360
Db |||||||
301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQDGNFEVWERPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVASLGKGVACNPACFIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Db |||||||
361 GPRSYTIAVASLGKGVACNPACFIITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
Db 421 M 421

```
RESULT 12
US-10-103-327-10
; Sequence 10, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-10

Query Match      100.0%; Score 2314; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAFORSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAFORSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYGYDIDAQTFADGWVLLKFDGVCYCDL 180
DB 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYGYDIDAQTFADGWVLLKFDGVCYCDL 180
QY 181 ADGYKHSALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKS 240
DB 181 ADGYKHSALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKS 240
QY 241 SILDWTSEFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQOMALWAIMAAPLFMS 300
DB 241 SILDWTSEFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQOMALWAIMAAPLFMS 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQ 360
DB 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQ 360
QY 361 GPRSYTIAVASLKGVCACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQ 420
DB 361 GPRSYTIAVASLKGVCACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQ 420
QY 421 M 421
DB 421 M 421

RESULT 13
US-09-993-059-6
; Sequence 6, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
```

```
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-6

Query Match      100.0%; Score 2314; DB 2; Length 435;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAFORSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAFORSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYGYDIDAQTFADGWVLLKFDGVCYCDL 180
DB 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGFYGYDIDAQTFADGWVLLKFDGVCYCDL 180
QY 181 ADGYKHSALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKS 240
DB 181 ADGYKHSALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKS 240
QY 241 SILDWTSEFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQOMALWAIMAAPLFMS 300
DB 241 SILDWTSEFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQOMALWAIMAAPLFMS 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQ 360
DB 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQ 360
QY 361 GPRSYTIAVASLKGVCACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQ 420
DB 361 GPRSYTIAVASLKGVCACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQ 420
QY 421 M 421
DB 421 M 421

RESULT 14
US-10-103-327-6
; Sequence 6, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-6

Query Match      100.0%; Score 2314; DB 2; Length 435;
Best Local Similarity 100.0%; Pred. No. 1.5e-234;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWFRCMCLDCQEEP 60
Db 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWFRCMCLDCQEEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQORDSEGRLOADPQRFPHGIROL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQORDSEGRLOADPQRFPHGIROL 120
QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGGYDYIDAQTFADWGVLLKFDGCYCDLLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGGYDYIDAQTFADWGVLLKFDGCYCDLLENL 180
QY 181 ADGYKXMSLALNRTGRSIVYSCWPLYMWPFPKPNYTEIROYCNHWRNFADIDDSWKSJK 240
Db 181 ADGYKXMSLALNRTGRSIVYSCWPLYMWPFPKPNYTEIROYCNHWRNFADIDDSWKSJK 240
QY 241 SILDWTSTFQNERIVDVAGPGWNDPDMVLVGNFGLSWNQVQVQTMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFQNERIVDVAGPGWNDPDMVLVGNFGLSWNQVQVQTMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
QY 421 M 421
Db 421 M 421
```

RESULT 15

```
US-09-070-356-4
; Sequence 4, Application US/09070356
; Patent No. 6228631
; GENERAL INFORMATION:
; APPLICANT: Alex Zhu
; APPLICANT: Jack Goldstein
; TITLE OF INVENTION: Recombinant a-N-
; TITLE OF INVENTION: Acetylglucosaminidase
; TITLE OF INVENTION: Enzyme and cDNA Encoding
; TITLE OF INVENTION: Said Enzyme
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,356
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,248
; FILING DATE: March 26, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pasqualini, Patricia A.
; REGISTRATION NUMBER: 34,894
; REFERENCE/DOCKET NUMBER: 63475/12
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 4:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: cDNA to mRNA
; HYPOTHETICAL: no
; ANTI-SENSE: yes
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: human
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE: library
; POSITION IN GENOME: unknown
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY: human a-galactosidase
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS: Calhoun et al
; TITLE: Fabry Disease: Isolation of a cDNA
; TITLE: Clone Encoding Human a-Galactosidase A
; JOURNAL: Proceedings of the National Academy
; JOURNAL: of Science USA
; VOLUME: 82
; PAGES: 7364-7368
; DATE: 1985
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
US-09-070-356-4
```

Query Match 99.8%; Score 2310; DB 2; Length 429;

Best Local Similarity 99.0%; Pred. No. 3.9e-234;

Matches 417; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWFRCMCLDCQEEP 60
Db 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWFRCMCLDCQEEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQORDSEGRLOADPQRFPHGIROL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQORDSEGRLOADPQRFPHGIROL 120
QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGGYDYIDAQTFADWGVLLKFDGCYCDLLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGGYDYIDAQTFADWGVLLKFDGCYCDLLENL 180
QY 181 ADGYKXMSLALNRTGRSIVYSCWPLYMWPFPKPNYTEIROYCNHWRNFADIDDSWKSJK 240
Db 181 ADGYKXMSLALNRTGRSIVYSCWPLYMWPFPKPNYTEIROYCNHWRNFADIDDSWKSJK 240
QY 241 SILDWTSTFQNERIVDVAGPGWNDPDMVLVGNFGLSWNQVQVQTMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFQNERIVDVAGPGWNDPDMVLVGNFGLSWNQVQVQTMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQETG 360
```


Qy	361	GPRSYTIAVASLGKGVACNPACFITQLLPVKRGKLGFEWTSRLRSHINPTGTVLLQLENT	420
Db	361	GPRSYTIAVASLGKGVACNPACFITQLLPVKRGKLGFEWTSRLRSHINPTGTVLLQLENT	420
Qy	421	M	421
Db	421	M	421

Search completed: December 24, 2005, 09:12:00
Job time : 45 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 24, 2005, 09:07:49 ; Search time 40 Seconds
(without alignments)
1012.681 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MQRNPELHGLCALALFLA.....RLRSHINPTGTVLLQLENTM 421
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2314	100.0	429	1 GBHUA	alpha-galactosidas
2	1811	78.3	419	2 JC4522	alpha-galactosidas
3	1051	45.4	405	2 S45522	alpha-N-acetylglala
4	952.5	41.2	411	2 A35485	alpha-N-acetylglala
5	930	40.2	358	2 A33265	alpha-N-acetylglala
6	862	37.3	451	2 T24018	hypothetical prote
7	716.5	31.0	434	2 T47748	alpha-galactosidas
8	712	30.8	378	2 T50781	alpha-galactosidas
9	687.5	29.7	422	2 T06388	alpha-galactosidas
10	682	29.5	425	2 T10860	alpha-galactosidas
11	681.5	29.5	411	2 S07472	alpha-galactosidas
12	676	29.2	680	2 T36472	probable secreted
13	609	26.3	436	2 T39118	probable alpha-gal
14	598	25.8	396	2 JC5558	alpha-galactosidas
15	557	24.1	469	2 S45453	alpha-galactosidas
16	543.5	23.5	545	2 S23582	alpha-galactosidas
17	521	22.5	471	2 S50312	alpha-galactosidas
18	519	22.4	471	2 S50311	alpha-galactosidas
19	516	22.3	444	2 S74221	alpha-galactosidas
20	514	22.2	471	2 JQ1021	alpha-galactosidas
21	510	22.0	471	2 S50310	alpha-galactosidas
22	502	21.7	471	1 GBBYAG	alpha-galactosidas
23	300.5	13.0	204	2 T04423	hypothetical prote
24	282.5	12.2	432	2 F83883	alpha-galactosidas
25	198	8.6	159	2 T04422	alpha-galactosidas
26	180.5	7.8	624	2 S74222	alpha-galactosidas
27	129	5.6	641	2 A55549	glucan 1,6-alpha-i
28	125	5.4	348	2 A70311	hypothetical prote
29	116	5.0	4199	2 S76412	hypothetical prote

ALIGNMENTS

RESULT 1

GBHUA

alpha-galactosidase (EC 3.2.1.22) A precursor - human
N:Alternate names: alpha-D-galactoside galactohydrolase; melibiase
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1987.#sequence revision 27-Oct-1995 #text_change 09-Jul-2004
C:Accession: S04081; A29608; A30214; S14879; A00896; B00896; I37140
R:Kornreich, R.; Desnick, R.J.; Bishop, D.F.
Nucleic Acids Res. 17, 3301-3302, 1989

A:Title: Nucleotide sequence of the human alpha-galactosidase A gene.
A:Reference number: S04081; MUID:89263745; PMID:2542896
A:Accession: S04081

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-429 <KOR>

A:Cross-references: UNIPROT:P06280; UNIPARC:UPI00000033A30; EMBL:X14448; NID:g31755; PIDN:R:Quinn, M.; Hantzopoulos, P.; Fidanza, V.; Calhoun, D.H.
Gene 58, 177-188, 1987

A:Title: A genomic clone containing the promoter for the gene encoding the human lysosom
A:Reference number: A29608; MUID:88112869; PMID:2892762
A:Accession: A29608

A:Molecule type: DNA

A:Residues: 1-64 <QUI>

A:Cross-references: UNIPARC:UPI000016A96A; GB:M18242; NID:g182944; PIDN:AAA52514.1; PID

R:Bishop, D.F.; Kornreich, R.; Desnick, R.J.

Proc. Natl. Acad. Sci. U.S.A. 85, 3903-3907, 1988

A:Title: Structural organization of the human alpha-galactosidase A gene: further eviden

A:Reference number: A30214; MUID:88234528; PMID:2836863

A:Accession: A30214

A:Molecule type: DNA

A:Residues: 1-64 <BIS>

A:Cross-references: UNIPARC:UPI000016A96A; EMBL:M20317; EMBL:J03249

R:Koide, T.; Ishiura, M.; Iwai, K.; Inoue, M.; Kaneda, Y.; Okada, Y.; Uchida, T.

FEBS Lett. 259, 353-356, 1990

A:Title: A case of Fabry's disease in a patient with no alpha-galactosidase A activity

A:Reference number: S14879; MUID:9002580; PMID:2152885

A:Accession: S14879

A:Molecule type: mRNA

A:Residues: 1, 'K', 3-39, 'S', 41-429 <KOI>

A:Cross-references: UNIPARC:UPI000011E285; EMBL:X16889

A:Experimental source: Fabry's disease patient

R:Bishop, D.F.; Calhoun, D.H.; Bernstein, H.S.; Hantzopoulos, P.; Quinn, M.; Desnick, R

Proc. Natl. Acad. Sci. U.S.A. 83, 4859-4863, 1986

A:Title: Human alpha-galactosidase A: nucleotide sequence of a cDNA clone encoding the

A:Reference number: A00896; MUID:86259694; PMID:3014515

A:Accession: A00896

A:Molecule type: mRNA

A:Residues: 27-429 <BT2>

A:Cross-references: UNIPARC:UPI0000000358; GB:M13571; NID:g178245; PIDN:AAA51676.1; PID

A:Experimental source: lung

A:Accession: B00896

A:Molecule type: protein

A;Residues: 32-55,'S',57-58,'R',60-65,'S',67-68;228-232,'N',234-237,'A',298-326;'L',334-
A;Cross-references: UNIPARC:UPI00001729C6; UNIPARC:UPI00001729C7; UNIPARC:UPI00001729C8;
R;Tsuiji, S.; Martin, B.M.; Kaslow, D.C.; Migeon, B.R.; Choudary, P.V.; Stubblefield, B.K
Eur. J. Biochem. 165, 275-280, 1987
A;Title: Signal sequence and DNA-mediated expression of human lysosomal alpha-galactosid
A;Reference number: 137140; MUID:87246603; PMID:3036505
A;Accession: J37140
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-429 <RES>
A;Cross-references: UNIPARC:UPI0000033A30; EMBL:X05790; NID:g28535; PIDN:CAA29232.1; PID
C;Genetics:
A;Gene: GDB:GLA
A;Cross-references: GDB:119272; OMIM:301500
A;Map position: Xq21.3-Xq22
A;Introns: 65/2; 123/3; 183/1; 213/3; 267/3; 333/3
C;Superfamily: alpha-galactosidase
C;Keywords: Fabry disease; glycolipid metabolism; glycoprotein; glycosidase; hydrolase;
F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-429/Product: alpha-galactosidase A #status predicted <MAT>
F;139,192,215,408/Binding site: carboxylate (Asn) (covalent) #status predicted
Query Match 100.0%; Score 2314; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.7e-190;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGLQADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGLQADPQRFPHGIRQL 120
QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYGYDIDAQTFADWGVLLKFDGCGYCDLSNL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYGYDIDAQTFADWGVLLKFDGCGYCDLSNL 180
QY 181 ADGYKHSALNRTGRSIVYSCWPLYMWPQKPNYTIROYCNHWRNPFADIDDSWSIK 240
DB 181 ADGYKHSALNRTGRSIVYSCWPLYMWPQKPNYTIROYCNHWRNPFADIDDSWSIK 240
QY 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVGNFGLSWNQVQTQMALWAIMAAPLWMSNDL 300
DB 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVGNFGLSWNQVQTQMALWAIMAAPLWMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYLRQGNFVWRPLSGLAWAVAMINRQIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYLRQGNFVWRPLSGLAWAVAMINRQIG 360
QY 361 GPRSYTIAVASLGKGVACNPAFCITQLLPVKRKLGFYEWTSRLSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNPAFCITQLLPVKRKLGFYEWTSRLSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421
RESULT 2
JC4522
alpha-galactosidase (EC 3.2.1.22) A precursor - mouse
N;Alternate names: alpha-D-galactoside galactohydrolase
C;Species: Mus musculus (house mouse)
C;Date: 08-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004
C;Accession: JC4522
R;Ohshima, T.; Murray, G.J.; Nagle, J.W.; Quirk, J.M.; Kraus, M.H.; Barton, N.W.; Brady,
Gene 166, 277-280, 1995
A;Title: Structural organization and expression of the mouse gene encoding alpha-galacto
A;Reference number: JC4522; MUID:96125203; PMID:8543175
A;Accession: JC4522
A;Molecule type: mRNA
A;Residues: 1-419 <OHS>

A;Cross-references: UNIPROT:P51569; UNIPARC:UPI00000018CA; GB:U34071; NID:g1141787; PIDN
A;Experimental source: kidney, C57BL
C;Comment: This enzyme is a lysosomal enzyme that hydrolyses the alpha-D-galactosyl resi
C;Genetics:
A;Gene: alpha Gal
A;Introns: 65/2; 123/3; 183/1; 213/3; 267/3; 333/3
C;Superfamily: alpha-galactosidase
C;Keywords: glycoprotein; glycosidase; hydrolase; lysosome
F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-419/Product: alpha-galactosidase A #status predicted <MAT>
F;139,192,215,408/Binding site: carboxylate (Asn) (covalent) #status predicted
Query Match 78.3%; Score 1811; DB 2; Length 419;
Best Local Similarity 78.2%; Pred. No. 2.3e-147;
Matches 327; Conservative 41; Mismatches 50; Indels 0; Gaps 0;
QY 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MKLLSRDTRLVCELALCPALVFWLSILGVRLDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGLQADPQRFPHGIRQL 120
DB 61 DACISEQLFMQAEMLVSDGWRDAGYDYLICIDDCWMAQRDSEGLQADPQRFPHGIRQL 120
QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYGYDIDAQTFADWGVLLKFDGCGYCDLSNL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYGYDIDAQTFADWGVLLKFDGCGYCDLSNL 180
QY 181 ADGYKHSALNRTGRSIVYSCWPLYMWPQKPNYTIROYCNHWRNPFADIDDSWSIK 240
DB 181 ENGYKYMALNRTGRSIVYSCWPLYMWPQKPNYTIROYCNHWRNPFADIDDSWSIK 240
QY 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVGNFGLSWNQVQTQMALWAIMAAPLWMSNDL 300
DB 241 NILSWTVYQKEIVEVAGPGGWNDDMLVGNFGLSWNQVQTQMALWAIMAAPLWMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYLRQGNFVWRPLSGLAWAVAMINRQIG 360
DB 301 RQISSQAKALLQDKDVIAINQDPLGKQGYCFRKNHIEVWRPLSNLAWAVARNLQIG 360
QY 361 GPRSYTIAVASLGKGVACNPAFCITQLLPVKRKLGFYEWTSRLSHINPTGTVLLQLE 419
DB 361 GPCPTYIQTSSIGRLGACNPGCIITQLLPVKRKLGFYEWTSRLSHINPTGTVLLQLE 419
RESULT 3
S45522
alpha-N-acetylgalactosaminidase - chicken
C;Species: Gallus gallus (chicken)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C;Accession: S45522; S43413
R;Davis, M.O.; Hata, J.; Smith, D.; Walker, J.C.
submitted to the EMBL Data Library, December 1993
A;Reference number: S45522
A;Accession: S45522
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-405 <DAV>
A;Cross-references: UNIPROT:Q90744; UNIPARC:UPI000000FC7C2; EMBL:L18754; NID:g435036; PID
R;Davis, M.O.; Hata, D.J.; Smith, D.; Walker, J.C.
Biochim. Biophys. Acta 1216, 296-298, 1993
A;Title: Cloning and sequence of a chicken alpha-N-acetylgalactosaminidase gene.
A;Reference number: S43413; MUID:94060104; PMID:8241271
A;Accession: S43413
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-161,'A',163-405 <DA2>
A;Cross-references: UNIPARC:UPI0000146A75; GB:L18754
C;Superfamily: alpha-galactosidase
Query Match 45.4%; Score 1051; DB 2; Length 405;
Best Local Similarity 53.0%; Pred. No. 3.2e-82;
Matches 206; Conservative 58; Mismatches 117; Indels 8; Gaps 6;

F;124,177,201/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match	40.2%	Score	930;	DB	2;	Length	358;
Best Local Similarity	57.7%	Pred.	No. 6.4e-72;				
Matches	176;	Conservative	40;	Mismatches	85;	Indels	4;
Gaps							
Qy	32	LDNGLARTPTMGWLHWRFCMLDCQEPRDSCISEKLFMEAEALMVSGWKADGAYEYLICI	91				
Dd	18	LDNGLLQTTPMGLAWERFCINCEDEPKNCISQLFMEWADRWAQGDWRDMGYLNI	77				
Qy	92	DDCWMAQRDSRGLQADPQRPFGHGIROLANYVHSKGLGIYADVGNKTCAAGPPG-SFG	150				
Dd	78	DDCWIG-GRDASRLMPDPKRFPHGIPELAYVHSLGLGIYADMGNFTCMGYPGTILD	136				
Qy	151	YYIDTAOTPADWGVDLLKPDGCYCDSLENLADGYKHMSLANRTORSIYVSCWPLYMWP	210				
Dd	137	KVVQDAQTFAEWKMDMLKDGF-STPEAAQGYPKMAAGLNATGRPFAFSCSWPAYEGG	195				
Qy	211	F-QKPNTETIRQYNHNWFENADIIDSWSKS IKSLDLWTSTFNQERI VDVVAGPGWNDDMLV	269				
Dd	196	LPPRVNYSLQADI CNLMRYNDIQDSWSKSVLSILNWFEHQDILQFVAGPGHNNDDMLL	255				
Qy	270	IGNFGLSNQQOVTQMALWAIMAAPLEFMSENDRHRISPPAKALLQDKDVIAINODPLGKKQY	329				
Dd	256	IGNFGLSEQSRAQMALWTVLAAAPLLMTDTLTITSAQNWDILQNLPMIKINQDPLIGIOGR	315				
Qy	330	QLROG 334					
Dd	316	RIRKG 320					

RESULT 6
T24018
hypothetical protein R07B7.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T24018
R:Harris, B.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z19830
A:Accession: T24018
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-451 <WIL>
A:Cross-references: UNIPROT:Q21801; UNIPARC:UPI000007F870; EMBL:Z75955; PIDN:CAB00120.1;
A:Experimental source: clone R07B7

Query Match	37.3%;	Score 862;	DB 2;	Length 451;
Best Local Similarity	43.6%;	Pred. NO. 5.8e-66;		
Matches	174;	Conservative 62;	Mismatches 149;	Indels 14; Gaps 6
Qy	28	GARALDNGLARPTPTMGWLHWRFCMNCLOQSEPDSCISEKLFMEAEMLVMSGWDAGYE	87	
Db	12	GAFCLDNLGLRTPPKMGWSWTAFCYEIDCVKHPTCCINEQLYKDWDQLVSGGYDKVGK	71	
Qy	88	YLCTDDCWMAQRORSEGRQLADPQRPPHGIIRQLANYVHSGKLGIGYADVGNKTCAGPPG	147	
Db	72	SVHIDDCWSEMERDSHGILVANKTRFPSPGMKALAKYMDHRGLKFGIYEDYGTCTCGGYPG	131	
Qy	148	SFGYYDIDAQTFADWGVDLLKFDGCGYCDLSLENLADGYKHMSLALNRTGRSIYVSCWPLY	207	
Db	132	SYKHEKVDAQTFAAWDVDYLKLDGNCIDQ-AWMPIGYPLFEKELNETGRPIWYSCSWAY	190	
Qy	208	MWPQKP---NYTEIRQYCNHWRNFADTDDSWKSIKSLDWTSTFNQERIVDVAGGGWND	264	
Db	191	L--IDHPELVNNYLIGKCYNTWRNPDDTINSWKSIIISIIYYDKNQDXXHIPHFGSGKXHD	248	
Qy	265	PDMLVIGNFLGSWNQCVTQMALWAIMAAPLFMSNDLRHISPAKALLQDKDVIAINQDPL	324	

```

Db      249  PDMVLVGNKGITLDMISIQFTVVCWISAPLMSNDLRIIGDSFKDVLKNKGAIKINQDPL 308
Qy      325  GKQGYQLRGDGNFVWER---PLSG--LAWAVAMINRQEIIGPSRYTIVASLKGKVACN 379
Db      309  GIMGRLLKNSTDIGVYVKQITPSKGGKKSAFAVYLNRRNEKGYKRIEIQLASIG---LTD 365
Qy      380  PACFITQLLPVKRKLGPYEWTSRLRSRSHINPTGTVLLQLE 418
Db      366  PAGYYVHDHWSHVDLGLLRSGDSIVVSIAPAGSVFFRAD 404

RESULT 7
T47748
alpha-galactosidase-like protein - Arabidopsis thaliana
N;Alternate names: protein F18021.270
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C;Accession: T47748
R;Benes, V.; Wurmbach, E.; Drzonek, H.; Ansorge, W.; Mewes, H.W.; Rudd, S.; Ler
submitted to the Protein Sequence Database, April 2000
A;Reference number: 224474
A;Accession: T47748
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-434 <BEN>
A;Cross-references: UNIPROT:Q9LYL2; UNIPARC:UPI00000A62BB; EMBL:ATF18021
A;Experimental source: cultivar Columbia; BAC clone F18021
C;Genetics:
A;Map position: 3
A;Introns: 73/2; 93/1; 112/1; 123/3; 158/2; 182/3; 202/2; 224/2; 252/2; 274/1;
A;Note: F18021.270
C;Superfamily: alpha-galactosidase

```

Query Match	31.0%;	Score	716.5;	DB	2;	Length	434;
Best Local Similarity	42.4%;	Pred. No.	1.7e-53;				
Matches	153;	Conservative	54;	Mismatches	107;	Indels	47;
Gaps							
Qy	11	GCALALFLALVSDIIPG	ARALDNLGLARTPTMG	LHWERFMCNLD	QCBEPC	ISCEKLF	70
Db	41	GLVPKSFNSIYD	FSMYGRQLNNGLART	PQMGWNSWNFFACN	-----	INETVK	90
Qy	71	EMAEI	MYSEGWKQAGY	YLICDDCWMA	PQRDSRGLQAD	PORPHG	IRPOLANYHVS
Db	91	ETADAL	VSSGLADLGY	THVNIDDCW	SNLLRSEGLVP	HPHETFP	SGIKLLADYHVS
Qy	131	LGI	IVADVGNKTCAG	FPGSGFY	YDIDAOT	FEADGW	DLKFDG
Db	151	LGI	YSDAGVTC	VEHPGSL	FHEVDDAD	IFASG	VDYLYKDNCF
Qy	187	MSL	ALNRTGRS	IVYS-C	EWPL---	YMPFPQ	KPNYTEIRQYCN
Db	206	MRD	ALNATGRS	IPYSL	CEWGVD	DPALW-----	AKEVNSMR
Qy	243	LD-----	WTISF	NOER	IVDVAG	PGGWNDD	MLVIGNFGL
Db	257	ADL	NNKWAAY-----	AGPG	GWNDD	PDML	EIGNMG
Qy	299	DLR	HSPOAK	ALLD	KDVIAL	NOPLG	KQGYQLR---
Db	309	DVR	NWTAET	LEILS	KEII	AVNO	PLGVQGRKI
Qy	356	R	356				
Db	369	R	369				

```

RESULT 8
T50781
alpha-galactosidase (EC 3.2.1.22) [imported] - coffee
C;Species: Coffea arabica (coffee)
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
C;Accession: T50781

```

R.Zhu, A.; Goldstein, J.
Gene 140, 227-231, 1994
A:Title: Cloning and functional expression of a cDNA encoding coffee bean alpha-galactosidase
A:Reference number: Z52235; MUID:94193002; PMID:8144030
A:Accession: T50781
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-378 <ZHU>
A:Cross-references: UNIPROT:Q42656; UNIPARC:UPI00001256A8; EMBL:L27992; PIDN:AAA33022.1
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase

Query Match 30.8%; Score 712; DB 2; Length 378;
Best Local Similarity 42.5%; Pred. No. 3.4e-53;
Matches 152; Conservative 51; Mismatches 113; Indels 42; Gaps 8;
QY 32 LDNGLARTPTMGLHWRFCMNLDCQEPDSCISEKLFMEAEMLVSEGWKADAGYEYLCI 91
DB 16 LANGLGLTPPMGWSNNHFRCLD-----EKLIRETADAMVSKGLAALGVYKYNL 65
QY 92 DDCWAPQDSEGRLOADPQRPFGHGIQRLANYVHSGKLGKIYADVGNKTC-A-GPPGSFG 150
DB 66 DDCWAEALNDSQNLVPRGSTPFGIKALADYVHSGKLGKIYSDAGTQTCSTKTPGSLG 125
QY 151 YYDIDAOTFADGWDLKFDGCGYCDSELENLADGYKHMSLALNRTORSIVYS-CEWPLYMW 209
DB 126 HEEQDAKTFASGVDYLYKDYCNENNI-SPKERYPIMSKALINSGRSIFPSLCEWNG--- 180
QY 210 PFQKNYETIRQYCNHWRNFADIDDSWKSIIKILD----WTSFNQERIVDVAGPGGWNDD 265
DB 181 --EEDPATWAKEVGVNSWRTTGDIDDSWSSMTSRADNDKWASY-----AGPGGWNDD 230
QY 266 DMLVIGFGLSNQVQTMALWMAAPLFMSNDRHISPOAKALLODKVIAINQDPLG 325
DB 231 DMLVGVNGGMMTTEYSRHSFWALAKAPLLIGCDIRSDMGATFQLLSNAEVIANNQDKLG 290
QY 326 KQGYQLRGQDNFEVWERPLSGLAWAVAMINR-----QETGGPRSYTIAVASL 372
DB 291 VQGNKVTKYGDLEWVAGPLSGKRVAVALNWRSGSTATITAYMSDVGLPSTAVVARNDL 348

RESULT 9
T06388
alpha-galactosidase (EC 3.2.1.22) - soybean
C:Species: Glycine max (soybean)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
C:Accession: T06388
R:Davis, M.O.; Walker, J.C.; Smith, D.
submitted to the EMBL Data Library, August 1994
A:Description: Cloning and expression of a soybean alpha galactosidase gene.
A:Reference number: Z15645
A:Accession: T06388
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-422 <DAV>
A:Cross-references: UNIPROT:Q39811; UNIPARC:UPI00000A7200; EMBL:U12926; PID:G927574; PID
A:Experimental source: strain williams
C:Function:
A:Description: catalyzes hydrolysis of melibiose into galactose and glucose
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase

Query Match 29.7%; Score 687.5; DB 2; Length 422;
Best Local Similarity 43.2%; Pred. No. 4.9e-51;
Matches 143; Conservative 45; Mismatches 112; Indels 31; Gaps 7;
QY 32 LDNGLARTPTMGLHWRFCMNLDCQEPDSCISEKLFMEAEMLVSEGWKADAGYEYLCI 91
DB 60 LDNGLHTPPMGWSNNHFACN-----IKEDLIRETADAMVSTGLAALGVYQYINI 109
QY 92 DDCWAPQDSEGRLOADPQRPFGHGIQRLANYVHSGKLGKIYADVGNKTC-A-GPPGSFG 150
DB 110 DDCWAEALNDSQNLVPRGSTPFGIKALADYVHSGKLGKIYSDAGTQTCSTKTPGSLG 169

QY 151 YYDIDAOTFADGWDLKFDGCGYCDSELENLADGYKHMSLALNRTORSIVYS-CEWPLYMW 209
DB 170 HEEQDAKTFASGVDYLYKDYCNENNI-SPKERYPIMSEALANTGRPIFFSLCEWNG---- 224
QY 210 PFQKNYETIRQYCNHWRNFADIDDSWKSIIKILD----WTSFNQERIVDVAGPGGWNDD 265
DB 225 --SDPATWAKSVGNSWRTTGDIDQKWDMSIRADLNDKWASY-----AGPGGWNDD 274
QY 266 DMLVIGFGLSNQVQTMALWMAAPLFMSNDRHISPOAKALLODKVIAINQDPLG 325
DB 275 DMLVGVNGGMMTTEYSRAHFSIWSLAKAPLLIGCDIRALDATTKEILLSKEVIANNQDKLG 334
QY 326 KQGYQLRGQDNFEVWERPLSGLAWAVAMINR 356
DB 335 VQGNKVTKYGDLEWVAGPLSGKRVAVILNWR 365

RESULT 10
T10860
alpha-galactosidase (EC 3.2.1.22) - kidney bean
C:Species: Phaseolus vulgaris (kidney bean)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T10860
R:Davis, M.O.; Walker, J.C.; Smith, D.
submitted to the EMBL Data Library, August 1994
A:Description: Cloning and expression of a pinto bean alpha galactosidase gene.
A:Reference number: Z17189
A:Accession: T10860
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-425 <DAV>
A:Cross-references: UNIPROT:Q41100; UNIPARC:UPI00000A0F13; EMBL:U12927; PID:G927576; PID
C:Function:
A:Description: catalyzes hydrolysis of melibiose into galactose and glucose
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase

Query Match 29.5%; Score 682; DB 2; Length 425;
Best Local Similarity 41.5%; Pred. No. 1.5e-50;
Matches 144; Conservative 54; Mismatches 113; Indels 36; Gaps 9;
QY 19 LALVSWDIPGARAL-DNGLARTPTMGLHWRFCMNLDCQEPDSCISEKLFMEAEMLV 77
DB 49 MMMSREVDHRRNLVGNLGTQTPPMGWSNNHFCN-----INEDLIRETADAMV 98
QY 78 SEGKWDAGVEYLCIDDCWAPQDSEGRLOADPQRPFGHGIQRLANYVHSGKLGKIYADV 137
DB 99 STGLAALGVYQYINIDDCWAEALNDSQNLVPRGSTPFGIKALADYVHSGKLGKIYSDA 158
QY 138 GNKTCA-GPPGSFGYDIDAOTFADGWVDLLKFDGCGYCDSELENLA--DGYKHMSLALNRT 194
DB 159 GTQTCSTKTPGSLGHEEQDAKTFASGIDYLYKDYCN----ENKNISPKERYPPMSKALANS 215
QY 195 GRSIVYS-CEWPLYMWPPQKNYETIRQYCNHWRNFADIDDSWKSIIKILD----WTSFN 249
DB 216 GRPIFFSLCEWNG-----SEDPATWAKSVGNSWRTTGDIEDKWSMISRADLNDWASY- 268
QY 250 QERIVDVAGPGGWNDDPDMLVIGNGLSNQVQTMALWMAAPLFMSNDRHISPOAKA 309
DB 269 -----AGPGGWNDDPDMLEVNGGMMTTEYSRAHFSIWSLAKAPLLIGCDIRALDVTYKE 321
QY 310 LLQDKDVIAINQDPLGKQGYQLRGQDNFEVWERPLSGLAWAVAMINR 356
DB 322 LLSNEEVIANNQDKLGQGVKKVKNNDLEWVAGPLSGNNRLAVILNWR 368

RESULT 11
S07472
alpha-galactosidase (EC 3.2.1.22) precursor - guar
C:Species: Cyanopsis tetragonoloba (guar, cluster bean)
C:Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 09-Jul-2004
C:Accession: S07472

R; Overbeeke, N.; Fellinger, A.J.; Toonen, M.Y.; van Wassenaar, D.; Verrips, C.T. *Plant Mol. Biol.* 13, 541-550, 1989

A;Accession: SU/4/2
A;Molecule type: mRNA
A;Residues: 1-411 <OVE>
A;Cross-references: UNIPROT:P14749; UNIPARC:UPI00001256A9; EMBL:X14619; NID:gl8291; PIDN
A;Note: the authors translated the codon ATT for residue 20 as Asn, TAT for residue 140
C;Superfamily: alpha-galactosidase
C;Keywords: glycosidase; hydrolase

Query Match	29.5%;	Score 681.5;	DB 2;	Length 411;
Best Local Similarity	42.2%;	Pred. No. 1.5e-50;		
Matches 141: Conservative	52;	Mismatches 102;	Indels 39;	Gaps 9;

Qy

33 DNLARTPMGWLHWERMCNLDQQEEDPDCISEKLFMEWAELMVSEGWKDAGYEYLCID 92
:
:

93 DCWMAPOBDSGRLOADPQRFPHGSIROLANYVHSHKGLKGIYADVGNKTKCA-GPGSGSGY 151

Db 99 DCWAELNRDSEGNMVPNAAAFPSGIGKALADYVHSGKLGWYSDAGNQTCRKMPGSLGH 158

Q7	101DAQITADWGVUJATDGLICUSHNTA	-----	DGINTSTJALNKIGNSIVIS	CENEF	200
D6	159 EEOAKTASGWVDYJKTQNC	-----	ENLGISVXERYPMGKALL	SSGRPIFFSCE	210

Qy 207 YMPFQKPNYTEIRQYCNHRNFADIDDSWKSIXSILD---WTSFNQERIVDVAGPGGW 262

DD	211	--GWGWEUFQ1W--AKASIGNSWK11GULEJUNWNSMISLAUSNDAMWAS1-----AGFGGW	233
QV	263	NBDPMLVIGNEGI:SWN00VYOTOMALWA1MAAPLEMSNDLRHISPOAKALLODKOVIAINOD	322

Db 260 NDFPMLEVGNGGMTTEERSHFSWALAKAPLLVGCDIRAMDDTHLHLSNAEVIANQD 319

QY 323 FLKGGIYQKQGFNFVMEKFLPSGLAWAVAMINK 356

Dp 320 KLGVOGKKYKSTNDLEWAGCPTSDNKQAVITLWNR 353

RESULT 12

probable secreted alpha-galactosidase - *Streptomyces coelicolor*
C. Species. *Streptomyces coelicolor*
1364/2

C:Accession: T36472
C:date: 03-dec-1999 #sequence_revision 03-dec-1999 #text_change 09-jul-2004
P:Seener K J : Harris D : Thomson N P : Barthill J : Barrell J : Barrell B G : Rajand

A; Reference number: Z21607
submitted to the EMBL Data Library, September 1999
Accession. T36472

A; Status: preliminary; translated from GB/EMBL/DBJ
A; Molecule type: DNA
A; Residues: 1,690

A; Cross-references: UNIPROT:Q9S2C9; UNIPARC:UPI00000B403; ENBL:AL110470; PIDN:CAB54163.
A; Experimental source: strain A3 (2)

A; Gene: SCOENB:SCP85.12

Best Local Similarity 38.2%; Pred. No. 8.9e-50;
Matches 161; Conservative 67; Mismatches 140; Indels 54; Gaps 14

[illegible]

71 EMAELMVSEGWKDAYEYLCLDDCWMAPORDSEGRQLQADPQRPFGIRQLANYVHSGKGLK 130

Db 81 GIADLFVERGLKDAGYEYVNLDDCWALPERDADGRLVPDPKFRFNGIEAVADYVHSGKGL 140

[illegible]

141	FGIYTSAGTKTCSGIFPGGALGHEYS	DARQFADGWVDYLYKYDNCNNQGV	D-AKQRYTTMR	199
189	LALNRTGRSIVYS-CEW----	PLYMPPQKPNYTEIRQYCNHNRNFAD	IDDWSKSIKSL	243
200	DALAATGRPIVYSICEWGENKP----	WEGA-----GDLGQL-----	WRTTGDINDSWGSSMSSIM	249
244	DWTSFNQERIVDVAGPGGWNDDP	MLVTGNFGLSNQGVQTMALWAIMAAP	FMSNDLRI	303
250	K-----SNLEIAEYARPGGWNDDP	MLEVNGGNTDEYRTHPSMHSIMAAP	LLITGDLRTA	305
304	SPOAKALLQDKDVIAINQDPLGKGYQ	LRQGDNFVMEERPUSGLAWAVAMIN	ROBIGPR	363
306	PEGAPELITINDEVIAVDQDPLGKQGV	SVSSGGRWVSKELADGSRVALFNE	-----GSR	361
364	SYTIAVASLKGKGVACNPACFITQLL	PKVRKLGFEYEW-----TSRLRSH	NPNTGTVLL	415
362	AQRIETAKAVG-----	LPKSRGYTMRDLWKHSDTNTTGRIA	AVPAGHTVLV	409
416	QL	417		
410	RV	411		

RESULT 13

T39118
Probable alpha-galactosidase - fission yeast (*Schizosaccharomyces pombe*)
C:Species: *Schizosaccharomyces pombe*
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T39118
R:Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, November 1999
A:Reference number: Z21829
A:Accession: T39118
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-436 <HUN>
A:Cross-references: UNIPROT:Q9URZ0; UNIPARC:UPI000006B88C; EMBL:AL132779; PIDN
A:Experimental source: strain 972H-; cosmid c869
C:Genetics:
A:Gene: SPDB:SPAC869.07C
A:Map position: 1
C:Superfamily: alpha-galactosidase

Query Match	26.3%;	Score 609;	DB 2;	Length 436;
Best Local Similarity	35.8%;	Pred. No. 2.7e-44;		
Matches 136: Conservative	65;	Mismatches 119;	Indels 60;	Gaps 14

Qy 10 LGCALALRFLALYSWDIPGARALDNLGLARTPTMGWLHWRFMCNLDQEEPDSCISEKLF 69

DD / LNC-FFJVLFFJFFSDVHSGI-1-1-NOTGKRFQWGNHNSMNAIACDID-1011 3
QV 70 MEMAELWYSEGWTDAGYEVLCIDCWMAEORD-SEGRLOADPQEPFHGFIQLANVYHSG 128

Db 53 LNNAKAIKEEGLLDGYEIVMDDCWSCXKHNATTTGRLEANPDKFPNGICSMAKKLHDMG 112

[illegible]

QY 185 KMSLALNRTGRSIVYS-CEW-PLYMWPQKPNYTEIRQYCNHWRNFADIDSWK----- 237

DB	173	KRMSDALNKGRIPIFYSLCQWGEDFVNWNG-----NIIANSWKISGJIFDIFSRKNVNR	222
Qv7	238	-----SIKSLIIMTSENQRIIVDVAG--PGGWNDDMLVIGNEGII--SNVQQ	280

Db 226 CPCETIECFALQGDHCSVMNIIKASFLSSK----AGMNSGWNLDLSLEFVNGGMSFEY 281

09	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
281	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
282	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
283	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
284	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
285	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
286	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
287	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
288	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
289	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
290	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
291	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
292	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
293	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
294	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
295	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
296	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
297	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
298	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
299	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333
300	VTQWALWALMAAPLFMSNDLRLHSFQAKALLQKDVIALNQDDEBNQ-----IQURQGDN	333

QY 337 F-EVWERPLSGLAWAVAMIN 355

Db 341 YIELFSGRLSNDDWVAVLN 360

RESULT 14

JC5558
alpha-galactosidase (EC 3.2.1.22) II precursor - Mortierella vinacea
N:Alternate names: melibiase
C:Species: Mortierella vinacea
C>Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 09-Jul-2004
C:Accession: JC5558; PC4479
R:Shibuya, H.; Kobayashi, H.; Sato, T.; Kim, W.S.; Yoshida, S.; Kaneko, S.; Kasano, K.;
Biosci. Biotechnol. Biochem. 61, 592-598, 1997
A:Title: Purification, characterization, and cDNA cloning of a novel alpha-galactosidase
A:Reference number: JC5558; MUID:97290877; PMID:9145516
A:Accession: JC5558
A:Molecule type: mRNA
A:Residues: 1-396 <SH1>
A:Cross-references: UNIPROT:Q93816; UNIPARC:UPI000006AE03; GB:AB018691; NID:g3777480; PI
A:Accession: PC4479
A:Molecule type: protein
A:Residues: 21-39; 178-187 <SH2>
A:Cross-references: UNIPARC:UPI0000175B1F; UNIPARC:UPI0000175B20
A:Note: the authors translated the codon TCC for residue 112 as Leu, ACC for residue 122
e 236 as Gln
A:Note: the authors translated the codon ATG for residue 268 as Asn, TCC for residue 310
C:Comment: This enzyme liberates both side-chain and terminal alpha-galactosyl residues
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-396/Product: alpha-galactosidase II #status predicted <MAT>

Query Match 25.8%; Score 598; DB 2; Length 396;
Best Local Similarity 37.4%; Pred. No. 2.1e-43;
Matches 126; Conservative 57; Mismatches 116; Indels 38; Gaps 9;

Qy 32 LQNGARTPTMGWLHWFRCNLDCEEPDSCISEKLFMEAEIYVSECKDAGYEYLCI 91
Db 22 IDPSLAKTQMGNSNKKYQCN-----VNVTIINTANMVSSGLKDLGVHYINI 71
Qy 92 DDCWMAQPD-SEGRLOADPQRPFGHGIROLANYVHSGKLGKIYADVGNKTCAGPFGSFG 150
Db 72 DDCWSLHQDNTQTQTAPDTPFNGISGVASKVHALGKFGIYSDAGTNTCAGYFGSYG 131
Qy 151 YVDIDQTFADWGVLLKFDGVCYDSENL-----ADGYKMSLALNTRGRSIVYS 201
Db 132 YEADAQAQFSDWGVYLYKYNDC-----NNLGLAGNATISSKRYKRGMDALKNVSRPIFPS 186
Qy 202 -CEWPL-YMPPQKPNYTEIRQYCNHWRNPFADIDDSWKSISILDWTSFNQERIVDVAGP 259
Db 187 LCSWGTDDVMDGRSTGGQ-----SWRMGDISDNWSSVSVI---TCQAVPIANISAP 236
Qy 260 GGWNDPDLVLVIG-NFGLSNWQVQTOMALWMAAPLFMSNDRHISPOAKALLQDKDVIA 318
Db 237 GGWNDMDLVEVGHQDQMTITETSHPSIWAAMKSPILINDITNMINDIKNIITNEVIA 296
Qy 319 INQDPLGKQGYQLRQGDNFNFWERPLSLGLAWAVAMIN 355
Db 297 ISQDSLGSVQQRSMKGNTOQLFAGPLSKNGYVSLFLN 333

RESULT 15

S45453
alpha-galactosidase (EC 3.2.1.22) MEL precursor - yeast (Zygosaccharomyces cidri)
N:Alternate names: MEL protein
C:Species: Zygosaccharomyces cidri
C>Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004
C:Accession: S45453
R:Turakainen, H.; Hankaanpaae, M.; Korhola, M.; Aho, S.
Yeast 10, 733-745, 1994
A:Title: Characterization of MEL genes in the genus Zygosaccharomyces.
A:Reference number: S45453; MUID:95066377; PMID:7975892
A:Accession: S45453
A:Molecule type: DNA

A:Residues: 1-469 <TUR>
A:Cross-references: UNIPROT:Q99172; UNIPARC:UPI00000696F1; EMBL:L24957; NID:g538518; PI
C:Genetics:
A:Gene: MEL
C:Superfamily: alpha-galactosidase
C:Keywords: glycoprotein; glycosidase; hydrolase
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-469/Product: alpha-galactosidase MEL #status predicted <MAT>

Query Match 24.1%; Score 557; DB 2; Length 469;
Best Local Similarity 35.3%; Pred. No. 8.7e-40;
Matches 134; Conservative 52; Mismatches 128; Indels 66; Gaps 12;

Qy 18 FLALV--SWDIPGALDNLGLARTPTMGWLHWFRCNLDCEEPDSCISEKLFMEAEI 75
Db 5 FFALFFSSTDVLAASPSVNLGLTQMGWNNNSFGCS-----VKEELLGLTAK 54
Qy 76 MYSEGKADAGYEYLCIDDCWMAQPDSEGRLOADPQRPFGHGIROLANYVHSGKLGKIYI 135
Db 55 IVKLGKDLGYNYIILDDCW-SSGRSSNGSLDADSKFPFGMKYVAEQJHNSQLKFGMYS 113
Qy 136 DVGNTKTCAGPFGSFGYDIDDAQTTFADWGVLLKFDGCV-----CDSLENLADGYKHMSLAL 191
Db 114 SAGEYTCAGYAGSLGYEDMDATFASWDVLYKYDNCYNKGEFGTPEISYKRYKMSDAL 173
Qy 192 NTRGRSIVYS-CEWP---LYMPPQKPNYTEIRQYCNHWRNPFADII----- 232
Db 174 NKTGRPIFYSLCNWQDLTFYWG-----SAISNSWRMSGDVVYPQFDRPDSRCPGSG 224
Qy 233 ---DDSWK-----SISILDWTSFNQERIVDVAGPFGWNPDMIVGNFGLSNQVQTQWA 285
Db 225 DEYDCSYPGFHCISIMNLNKAAPMGON-----AAPGWNLDMLLEVGVGNMSDSEVAHFS 280
Qy 286 LWAIMAAPLFMSNDRHISPOAKALLQDKDVIAINQDPLGKQ-----YQLRQGDNF--- 337
Db 281 MWAIYKSPILIGADIDDLKDSLSVSNPAVIAINQDVLGTPTATRIWKYHVSQKQYGG 340
Qy 338 --EYWERPLSLGLAWAVAMIN 355
Db 341 ETQLWSGPLDNGDHVVALLN 360

Search completed: December 24, 2005, 09:12:46
Job time : 42 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1266	100.0	1266	6	AR653696	AR653696	Sequence
2	1266	100.0	1266	6	AR654964	AR654964	Sequence
3	1264	99.8	1284	6	AR653697	AR653697	Sequence
4	1264	99.8	1284	6	AR654965	AR654965	Sequence
5	1263.4	99.8	1278	6	AR653694	AR653694	Sequence
6	1263.4	99.8	1278	6	AR654962	AR654962	Sequence
7	1263.4	99.8	1290	6	AR653692	AR653692	Sequence
8	1263.4	99.8	1290	6	AR654960	AR654960	Sequence
9	1263.4	99.8	1296	6	AR653695	AR653695	Sequence
10	1263.4	99.8	1296	6	AR654963	AR654963	Sequence
11	1263.4	99.8	1308	6	AR653693	AR653693	Sequence
12	1263.4	99.8	1308	6	AR654961	AR654961	Sequence
13	1261.8	99.7	1290	6	CS135670	CS135670	Sequence
14	1261.8	99.7	1290	8	BT006884	BT006884	Homo sapi
15	1261.8	99.7	1290	11	AY889631	AY889631	Syntheti
16	1261.8	99.7	1290	11	AY889632	AY889632	Syntheti
17	1261.8	99.7	1290	11	AY892090	AY892090	Syntheti
18	1261.8	99.7	1290	11	BT007835	BT007835	Syntheti

241 TGGAGGATGCGAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAAGA 300
Db |||||||
241 TGGAGGATGCGAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAAGA 300
QY |||||||
301 GATTTCAGAAAGGAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db |||||||
301 GATTTCAGAAAGGAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY |||||||
361 GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAATAAA 420
Db |||||||
361 GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAATAAA 420
QY |||||||
421 ACTGCGGAGGCTTCCTCGGAGTTTGGATACACTACGACATTCATGATGCTCCAGACCTTTGCT 480
Db |||||||
421 ACTGCGGAGGCTTCCTCGGAGTTTGGATACACTACGACATTCATGATGCTCCAGACCTTTGCT 480
QY |||||||
481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Db |||||||
481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
QY |||||||
541 GCAGATGGTTATAAGCACATGCTTGGCCCTGAATAGGACTGGCGAAGCAATTGTGTAC 600
Db |||||||
541 GCAGATGGTTATAAGCACATGCTTGGCCCTGAATAGGACTGGCGAAGCAATTGTGTAC 600
QY |||||||
601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
Db |||||||
601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
QY |||||||
661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTCATTCATTCCTGGAAAAGTATAAG 720
Db |||||||
661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTCATTCATTCCTGGAAAAGTATAAG 720
QY |||||||
721 AGTATCTTGGACTGACATCTTTTAACAGGAGAGAAATTTGTGATGTTGCTGACAGGG 780
Db |||||||
721 AGTATCTTGGACTGACATCTTTTAACAGGAGAGAAATTTGTGATGTTGCTGACAGGG 780
QY |||||||
781 GGTGGAAATGACCCAGATGTTAGTATGTCGCACTTTGGCTCAGCTGGAAATCAGCAA 840
Db |||||||
781 GGTGGAAATGACCCAGATGTTAGTATGTCGCACTTTGGCTCAGCTGGAAATCAGCAA 840
QY |||||||
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
Db |||||||
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
QY |||||||
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCGCATCAAT 960
Db |||||||
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCGCATCAAT 960
QY |||||||
961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Db |||||||
961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
QY |||||||
1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGATGCTTATGATAAACCGGAGGAGATTGCT 1080
Db |||||||
1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGATGCTTATGATAAACCGGAGGAGATTGCT 1080
QY |||||||
1081 GSACCTCGCTCTTATACCATCGAGTTGCTTCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
Db |||||||
1081 GSACCTCGCTCTTATACCATCGAGTTGCTTCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
QY |||||||
1141 GCCTGCTTCATCACACAGCTCTCCTCGTGAAGGAAAGCTAGGTTCTATCAATGGACT 1200
Db |||||||
1141 GCCTGCTTCATCACACAGCTCTCCTCGTGAAGGAAAGCTAGGTTCTATCAATGGACT 1200
QY |||||||
1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
Db |||||||
1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
QY |||||||
1261 ATGTAA 1266
Db |||||||
1261 ATGTAA 1266

RESULT 2
AB654964
LOCUS AR654964 1266 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 11 from patent US 6890748.
ACCESSION AR654964
VERSION AR654964.1 GI:67586503
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1266)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 11 10-MAY-2005;
FEATURES large Scale Biology Corporation; Vacaville, CA
location/Qualifiers
1..1266
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 100.0%; Score 1266; DB 6; Length 1266;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGAGGCTGGCGCTTGGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGAGGCTGGCGCTTGGCTTCCTGGCC 60
QY 61 CTCGTTTCTCTGGACATCCCTGGGCTAGAGCACTGGCAATGGATTCGCAAGGACGCT 120
Db 61 CTCGTTTCTCTGGACATCCCTGGGCTAGAGCACTGGCAATGGATTCGCAAGGACGCT 120
QY 121 ACCATGGGCTGCTGCACTGGAGGCGTTTCATGTGCAACCTTTGACTGCCAGGAAGGCA 180
Db 121 ACCATGGGCTGCTGCACTGGAGGCGTTTCATGTGCAACCTTTGACTGCCAGGAAGGCA 180
QY 181 GATTCTCCATCAGTGAGAACTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCTCCATCAGTGAGAACTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATTCATGATGCTGTTGATGGCTCCCAAAGA 300
Db 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATTCATGATGCTGTTGATGGCTCCCAAAGA 300
QY 301 GATTTCAGAAAGGAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAAAGGAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAATAAA 420
Db 361 GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAATAAA 420
QY 421 ACCTCGGAGGCTTCCTCGGAGTTTGGATACACTACGACATTCATGATGCTCCAGACCTTTGCT 480
Db 421 ACCTCGGAGGCTTCCTCGGAGTTTGGATACACTACGACATTCATGATGCTCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
QY 541 GCAGATGGTTATAAGCACATGCTTGGCCCTGAATAGGACTGGCGAAGCAATTGTGTAC 600
Db 541 GCAGATGGTTATAAGCACATGCTTGGCCCTGAATAGGACTGGCGAAGCAATTGTGTAC 600
QY 601 TCCTGTGAGTGGCCCTCTTTTATATGTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCCCTCTTTTATATGTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTCATTCATTCCTGGAAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTCATTCATTCCTGGAAAAGTATAAG 720

```
QY 721 AGTATCTGGAGTGCACATCTTTTAAACAGAGAGAAATGTTGATGTTGCTGGACAGGG 780
DB 721 AGTATCTTTGGAGTGCACATCTTTTAAACAGAGAGAAATGTTGATGTTGCTGGACAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTGAATGSCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
DB 781 GGTGGAAATGACCCAGATATGTTAGTGAATGSCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACATGAGTGGCCCTCTGGGCTATCATGCTGCTCCTTTTATTCATGCTTAATGACCTC 900
DB 841 GTAACATGAGTGGCCCTCTGGGCTATCATGCTGCTCCTTTTATTCATGCTTAATGACCTC 900
QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGATAAGGAGTAAATGCGCATCAAT 960
DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGATAAGGAGTAAATGCGCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGAGCAAACTTTGMAAGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGAGCAAACTTTGMAAGTGG 1020
QY 1021 GAAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAAACCGGAGAGATTGGT 1080
DB 1021 GAAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAAACCGGAGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATGCACTGCTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
DB 1081 GGACCTCGCTCTTATACCATGCACTGCTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCTCCCTGCTGAAGAGGAGCTAGGTTCTATGATGAGT 1200
DB 1141 GCCTGCTTCATCACACAGCTCTCCCTGCTGAAGAGGAGCTAGGTTCTATGATGAGT 1200
QY 1201 TCAAGTTTAAAGAGTACATAAAATCCACAGGCACTGTTTTCCTCAGCTAGAGAAACACA 1260
DB 1201 TCAAGTTTAAAGAGTACATAAAATCCACAGGCACTGTTTTCCTCAGCTAGAGAAACACA 1260
QY 1261 ATGTAA 1266
DB 1261 ATGTAA 1266

RESULT 3
LOCUS AR653697
DEFINITION Sequence 13 from patent US 6887696.
ACCESSION AR653697
VERSION AR653697.1 GI:67584282
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1284)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 13 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source
1..1284
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 99.8%; Score 1264; DB 6; Length 1284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTCGGGCTTGGCGCTTTCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTCGGGCTTGGCGCTTTCGCTTCCTGGCC 60
QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGATGGCAAGGAGCGCT 120
DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGATGGCAAGGAGCGCT 120
```

```
QY 121 ACCATGGGCTGGCTGCACATCTGGAGCGCTTCATGTGCAACCTTGACTGCGAGAGAGCCA 180
DB 121 ACCATGGGCTGGCTGCACATCTGGAGCGCTTCATGTGCAACCTTGACTGCGAGAGAGCCA 180
QY 181 GATTTCCTGCATCAGTGGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
DB 181 GATTTCCTGCATCAGTGGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
QY 241 TGGAAAGATGAGGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAGA 300
DB 241 TGGAAAGATGAGGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAGA 300
QY 301 GATTTCAGAGGAGCACTTCAGGAGAGCCCTCAGGCTTTTCTCATGGGATTCGCCAGCTA 360
DB 301 GATTTCAGAGGAGCACTTCAGGAGAGCCCTCAGGCTTTTCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTATGTTTACAGCAAGGAGTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
DB 361 GCTAATTATGTTTACAGCAAGGAGTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTTGATGCCAGACCTTTGCT 480
DB 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTTGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTGAATCTGCTAATAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
DB 481 GACTGGGAGTGAATCTGCTAATAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
QY 541 GCAGATGTTTATAGCACATGCTCTGGCCCTGATAGGACTGGCAGAGCAATTTGTGTAC 600
DB 541 GCAGATGTTTATAGCACATGCTCTGGCCCTGATAGGACTGGCAGAGCAATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAGAGCCCAATTTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAGAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTTGATGATTTCTGCGAAAGTATAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTTGATGATTTCTGCGAAAGTATAAG 720
QY 721 AGTATCTTGGAGTGCACATCTTTTAAACAGAGAGAAATGTTGATGTTGCTGGACAGGG 780
DB 721 AGTATCTTGGAGTGCACATCTTTTAAACAGAGAGAAATGTTGATGTTGCTGGACAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTGAATGSCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
DB 781 GGTGGAAATGACCCAGATATGTTAGTGAATGSCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACATGAGTGGCCCTCTGGGCTATCATGCTGCTCCTTTTATTCATGCTTAATGACCTC 900
DB 841 GTAACATGAGTGGCCCTCTGGGCTATCATGCTGCTCCTTTTATTCATGCTTAATGACCTC 900
QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGATAAGGAGTAAATGCGCATCAAT 960
DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGATAAGGAGTAAATGCGCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGAGCAAACTTTGMAAGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGAGCAAACTTTGMAAGTGG 1020
QY 1021 GAAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAAACCGGAGAGATTGGT 1080
DB 1021 GAAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAAACCGGAGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATGCACTGCTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
DB 1081 GGACCTCGCTCTTATACCATGCACTGCTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCTCCCTGCTGAAGAGGAGCTAGGTTCTATGATGAGT 1200
DB 1141 GCCTGCTTCATCACACAGCTCTCCCTGCTGAAGAGGAGCTAGGTTCTATGATGAGT 1200
```

```
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGACCTGTTTTCAGCTCAGTAAACACA 1260
|
|
|
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGACCTGTTTTCAGCTCAGTAAACACA 1260
|
|
|
QY 1261 ATGT 1264
|
|
|
Db 1261 ATGT 1264

RESULT 4
AR654965
LOCUS AR654965 1284 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 13 from patent US 6890748.
ACCESSION AR654965
VERSION AR654965.1 GI:67586504
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1284)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 13 10-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source
1..1284
/mol_type="genomic DNA"
ORIGIN
Query Match 99.8%; Score 1264; DB 6; Length 1284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTTCGCTTCTCGGCC 60
|
|
|
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTTCGCTTCTCGGCC 60
|
|
|
QY 61 CTCGTTTCTGGACATCCCTGGGCTAGAGCACTGGAATGGATGGCAAGAGCGCT 120
|
|
|
Db 61 CTCGTTTCTGGACATCCCTGGGCTAGAGCACTGGAATGGATGGCAAGAGCGCT 120
|
|
|
QY 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCACTTGTACTGCCAGGAGCA 180
|
|
|
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCACTTGTACTGCCAGGAGCA 180
|
|
|
QY 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
|
|
|
Db 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
|
|
|
QY 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
|
|
|
Db 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
|
|
|
QY 301 GATTCAGAAGGAGAGCTTCAGGACAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
|
|
|
Db 301 GATTCAGAAGGAGAGCTTCAGGACAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
|
|
|
QY 361 GCTAATATTGTTACAGCAAAAGGAGCTAGAGCTAGGAGTTTATGCAAGTGTGGAAATAAA 420
|
|
|
Db 361 GCTAATATTGTTACAGCAAAAGGAGCTAGAGCTAGGAGTTTATGCAAGTGTGGAAATAAA 420
|
|
|
QY 421 ACTCGGCAAGGCTTCCTGGGAGTTTGGATACGACATGATGATGATGATGATGATGATGATGAT 480
|
|
|
Db 421 ACTCGGCAAGGCTTCCTGGGAGTTTGGATACGACATGATGATGATGATGATGATGATGATGAT 480
|
|
|
QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTGTGTTACTGTGACAGTGTGGAAATTTG 540
|
|
|
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTGTGTTACTGTGACAGTGTGGAAATTTG 540
|
|
|
QY 541 GCAGATGGTTATAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTGTAC 600
|
|
|
Db 541 GCAGATGGTTATAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTGTAC 600
|
|
|
```

```
QY 601 TCCTGTGAGTGGCCCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
|
|
|
Db 601 TCCTGTGAGTGGCCCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
|
|
|
QY 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTCGAAAGATATAAG 720
|
|
|
Db 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTCGAAAGATATAAG 720
|
|
|
QY 721 AGTATCTTTGGAATGGAATCTTTTAAACAGAGAGAAATTTGATGATTTGCTGGACAGGG 780
|
|
|
Db 721 AGTATCTTTGGAATGGAATCTTTTAAACAGAGAGAAATTTGATGATTTGCTGGACAGGG 780
|
|
|
QY 781 GGTGGAAATGACCCAGATATGTTAGTGAATTTGGCACTTTGGCCCTCAGCTGGAATCAGAA 840
|
|
|
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATTTGGCACTTTGGCCCTCAGCTGGAATCAGAA 840
|
|
|
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCTTAATGACCTC 900
|
|
|
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCTTAATGACCTC 900
|
|
|
QY 901 CGACATCATGAGCCCTCAAGCCAAAGCTCTCTTCAAGATAAGGACGTAAATTTGCCATCAAT 960
|
|
|
Db 901 CGACATCATGAGCCCTCAAGCCAAAGCTCTCTTCAAGATAAGGACGTAAATTTGCCATCAAT 960
|
|
|
QY 961 CAGGACCCCTTTGGGCAAGCGGTACCAGCTTACAGAGGGGAGGAGCAACTTTGAAGTGTGG 1020
|
|
|
Db 961 CAGGACCCCTTTGGGCAAGCGGTACCAGCTTACAGAGGGGAGGAGCAACTTTGAAGTGTGG 1020
|
|
|
QY 1021 GAACGACCTCTCTCAGGCTTAGCCCTGAGCTGTAGCTATGATATAAACCAGCAGAGATTGGT 1080
|
|
|
Db 1021 GAACGACCTCTCTCAGGCTTAGCCCTGAGCTGTAGCTATGATATAAACCAGCAGAGATTGGT 1080
|
|
|
QY 1081 GGACCTCGCTCTTATACCATCCAGATTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
|
|
|
Db 1081 GGACCTCGCTCTTATACCATCCAGATTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
|
|
|
QY 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAGCTAGGTTCTATGATGAGTACT 1200
|
|
|
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAGCTAGGTTCTATGATGAGTACT 1200
|
|
|
QY 1201 TCAAGTTTAAAGAGTCACATAAATCCACAGGCACTGTTTTCAGCTAGAAAAACA 1260
|
|
|
Db 1201 TCAAGTTTAAAGAGTCACATAAATCCACAGGCACTGTTTTCAGCTAGAAAAACA 1260
|
|
|
QY 1261 ATGT 1264
|
|
|
Db 1261 ATGT 1264

RESULT 5
AR653694
LOCUS AR653694 1278 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 7 from patent US 6887696.
ACCESSION AR653694
VERSION AR653694.1 GI:67584279
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1278)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 7 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source
1..1278
/mol_type="genomic DNA"
ORIGIN
Query Match 99.8%; Score 1263.4; DB 6; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
```


Db	421	ACTGCGCAGGCTTCCTCGGAGTTTGGATACACGACATGATGCCAGACCTTTGCT	480
Qy	481	GACTGGGAGTAGACTCTGCTAAAATTTGATGGTTGTACTGTGACAGATTTGGAAAATTTG	540
Db	481	GACTGGGAGTAGACTCTGCTAAAATTTGATGGTTGTACTGTGACAGATTTGGAAAATTTG	540
Qy	541	GCAGATGGTTATAAGCACATGCTTTGGCCCTGTAATAGACTGGCAGAGCAATTTGTATC	600
Db	541	GCAGATGGTTATAAGCACATGCTTTGGCCCTGTAATAGACTGGCAGAGCAATTTGTATC	600
Qy	601	TCCTGTGAGTGGCCCTTTATATGTCGGCCCTTCAAAGCCCAATATATACAGAAATCCGA	660
Db	601	TCCTGTGAGTGGCCCTTTATATGTCGGCCCTTCAAAGCCCAATATATACAGAAATCCGA	660
Qy	661	CAGTACTGCAATCACATGGCGAAATTTTGCTGCATTTGATGATTTCTTGGAATAAGTATAAG	720
Db	661	CAGTACTGCAATCACATGGCGAAATTTTGCTGCATTTGATGATTTCTTGGAATAAGTATAAG	720
Qy	721	AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACCAGGG	780
Db	721	AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACCAGGG	780
Qy	781	GGTTGGAATGACCCAGATATTTAGTATGATTTGGCACTTTGGCCCTCAGCTGGAAATCAGAA	840
Db	781	GGTTGGAATGACCCAGATATTTAGTATGATTTGGCACTTTGGCCCTCAGCTGGAAATCAGAA	840
Qy	841	GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCAATGCTAATGACCTC	900
Db	841	GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCAATGCTAATGACCTC	900
Qy	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAAATTCGCATCAAT	960
Db	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAAATTCGCATCAAT	960
Qy	961	CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG	1020
Db	961	CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG	1020
Qy	1021	GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT	1080
Db	1021	GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT	1080
Qy	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTCGGGTAAAGGAGTGGCCCTGTAATCCT	1140
Db	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTCGGGTAAAGGAGTGGCCCTGTAATCCT	1140
Qy	1141	GCCTGCTTCATCAGACAGCTCCCTCTGTGAAGGAAAGCTAGGGTTCTATGAATGGACT	1200
Db	1141	GCCTGCTTCATCAGACAGCTCCCTCTGTGAAGGAAAGCTAGGGTTCTATGAATGGACT	1200
Qy	1201	TCAAAGCTTAAGAGTCACATAAATCCACAGGCACCTGTTTTCCTTCAGCTAGAAAACACA	1260
Db	1201	TCAAAGCTTAAGAGTCACATAAATCCACAGGCACCTGTTTTCCTTCAGCTAGAAAACACA	1260
Qy	1261	ATGTA 1265	
Db	1261	ATGCA 1265	

RESULT 7	
AR653692	
LOCUS	AR653692 1290 bp DNA linear PAT 13-JUN-2005
DEFINITION	Sequence 3 from patent US 6887696.
ACCESSION	AR653692
VERSION	AR653692.1
KEYWORDS	GI:67584277
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 1290)
TITLE	Garger, S.J., Turpen, T.H. and Kumagai, M.H.
JOURNAL	Production of lysosomal enzymes in plants by transient expression Patent: US 6887696-A 3 03-MAY-2005;

FEATURES		Large Scale Biology Corporation; Vacaville, CA	
source		Location/Qualifiers	
		1..1290	
		/organism="unknown"	
		/mol_type="genomic DNA"	
ORIGIN			
Query Match		99.8%; Score 1263.4; DB 6; Length 1290;	
Best Local Similarity		99.9%; Pred. No. 0;	
Matches 1264; Conservative		0; Mismatches 1; Indels 0; Gaps 0;	
Qy	1	ATGCAAGCTGAGGAACCCAGAACTACATCTGGGCTCGGGCTTGCCTTCCTTCTCGCC	60
Db	1	ATGCAAGCTGAGGAACCCAGAACTACATCTGGGCTCGGGCTTGCCTTCCTTCTCGCC	60
Qy	61	CTCGTTTCTCTGGACATCCCTGGGCTAGAGCACATGGAATGGAAGGAGCGCT	120
Db	61	CTCGTTTCTCTGGACATCCCTGGGCTAGAGCACATGGAATGGAAGGAGCGCT	120
Qy	121	ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGAGCCA	180
Db	121	ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGAGCCA	180
Qy	181	GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240
Db	181	GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240
Qy	241	TGGAGGATGCAAGTTATGATACCTCTGCAATGATGATGCTGTGATGCTGCTCCCAAGA	300
Db	241	TGGAGGATGCAAGTTATGATACCTCTGCAATGATGATGCTGTGATGCTGCTCCCAAGA	300
Qy	301	GATTCAGAAGGAGAGACTTCAGGCAGACCTTCAGCGCTTTCTCATGGGATTCGCCAGCTA	360
Db	301	GATTCAGAAGGAGAGACTTCAGGCAGACCTTCAGCGCTTTCTCATGGGATTCGCCAGCTA	360
Qy	361	GCTAATATGTTTCA CAGCAAGGACTGAAGCTAGGGAATTTATGAGATGTTTGGAAATAAA	420
Db	361	GCTAATATGTTTCA CAGCAAGGACTGAAGCTAGGGAATTTATGAGATGTTTGGAAATAAA	420
Qy	421	ACCTCGGAGGCTCCCTGGGAGTTTGGATACAGCATTCAGCATTCGCGACACCTTTGCT	480
Db	421	ACCTCGGAGGCTCCCTGGGAGTTTGGATACAGCATTCAGCATTCGCGACACCTTTGCT	480
Qy	481	GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTACTGTGACAGATTTGGAAAATTTG	540
Db	481	GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTACTGTGACAGATTTGGAAAATTTG	540
Qy	541	GCAGATGGTTATAAGCACATGCTTCCTGGCCCTGAATAGGACTGGCAGAGCAATTTGTATC	600
Db	541	GCAGATGGTTATAAGCACATGCTTCCTGGCCCTGAATAGGACTGGCAGAGCAATTTGTATC	600
Qy	601	TCCTGTGAGTGGCCCTTTATATGTCGGCCCTTCAAAGCCCAATATATACAGAAATCCGA	660
Db	601	TCCTGTGAGTGGCCCTTTATATGTCGGCCCTTCAAAGCCCAATATATACAGAAATCCGA	660
Qy	661	CAGTACTGCAATCACATGGGAAATTTTGCTGACATTTGATGATTCCTGGAAAAGTATAAG	720
Db	661	CAGTACTGCAATCACATGGGAAATTTTGCTGACATTTGATGATTCCTGGAAAAGTATAAG	720
Qy	721	AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACCAGGG	780
Db	721	AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACCAGGG	780
Qy	781	GGTTGGAATGACCCAGATATTTAGTATGATTTGGCACTTTGGCCCTCAGCTGGAAATCAGAA	840
Db	781	GGTTGGAATGACCCAGATATTTAGTATGATTTGGCACTTTGGCCCTCAGCTGGAAATCAGAA	840
Qy	841	GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCAATGCTAATGACCTC	900
Db	841	GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCAATGCTAATGACCTC	900
Qy	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAAATTCGCATCAAT	960

Db 901 CGACATCATGCGCCCTCAGCCAAAGCTCTCCTTACAGGATAAGGACGTAAATGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGAGTACAGCTTTAGACAGGAGACAACTTTGAAGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGAGTACAGCTTTAGACAGGAGACAACTTTGAAGTGG 1020
Qy 1021 GAACACCTCTCTCAGGCTTACGCTGGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
Db 1021 GAACACCTCTCTCAGGCTTACGCTGGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCTCCTCTGTGAAAGAAAGTACAGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCTCCTCTGTGAAAGAAAGTACAGGCTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
Qy 1261 ATGTA 1265
Db 1261 ATGCA 1265

RESULT 8
LOCUS AR654960
DEFINITION Sequence 3 from patent US 6890748.
ACCESSION AR654960
VERSION AR654960.1 GI:67586497
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 3 10-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
Location/Qualifiers
source 1..1290
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 99.8%; Score 1263.4; DB 6; Length 1290;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGGCGCTTGGCGCTTCTTGGCC 60
Db 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGGCGCTTGGCGCTTCTTGGCC 60
Qy 61 CTCGTTTCTCGGACATCCTCGGGCTAGACACTGGCAATGGATTTGGCAAGGACGCT 120
Db 61 CTCGTTTCTCGGACATCCTCGGGCTAGACACTGGCAATGGATTTGGCAAGGACGCT 120
Qy 121 ACCATGGGCTGCTGCTGAGCTGGAGGCTTTCATGTGCAACCTTTGACTGGCAGAGGCA 180
Db 121 ACCATGGGCTGCTGCTGAGCTGGAGGCTTTCATGTGCAACCTTTGACTGGCAGAGGCA 180
Qy 181 GATTCTGCTCATGCTGAGAGCTCTTCATGGAGTGGCAGAGCTCATGCTCTCAGAGGC 240
Db 181 GATTCTGCTCATGCTGAGAGCTCTTCATGGAGTGGCAGAGCTCATGCTCTCAGAGGC 240
Qy 241 TGGAGGATGCTAGGTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
Db 241 TGGAGGATGCTAGGTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
Qy 301 GATTTCAGAGGAGGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTTCATGGGATTCGCCAGCTA 360

Db 301 GATTTCAGAGGAGGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATATGTTTACAGCAAAAGGAGTAAAGCTAGGAGTTTATGCAAGATGTTTGGAAATAAA 420
Db 361 GCTAATATGTTTACAGCAAAAGGAGTAAAGCTAGGAGTTTATGCAAGATGTTTGGAAATAAA 420
Qy 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATATACGACATTTGATGCTGAGAGTTTGGAAATTTG 480
Db 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATATACGACATTTGATGCTGAGAGTTTGGAAATTTG 480
Qy 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 541 CGAGATGGTTTAAAGCACATGTCCTTGGCCCTGAATAGGAGTGGCAGAGCAATTTGTGTAC 600
Db 541 CGAGATGGTTTAAAGCACATGTCCTTGGCCCTGAATAGGAGTGGCAGAGCAATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCCCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCCCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCATCTACCTGGGAAATTTTCTGACATTTGATGATTTCTCTGAAAGATTAAG 720
Db 661 CAGTACTGCATCTACCTGGGAAATTTTCTGACATTTGATGATTTCTCTGAAAGATTAAG 720
Qy 721 AGTATCTTGGAGTGGACATCTTTTAAACAGAGAGAAATTTGTGTTGCTGGACAGGG 780
Db 721 AGTATCTTGGAGTGGACATCTTTTAAACAGAGAGAAATTTGTGTTGCTGGACAGGG 780
Qy 781 GGTGGAAATGACCCAGAGATATGTTAGTGTATGGCAATTTGGCCCTCAGCTGGAAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGAGATATGTTAGTGTATGGCAATTTGGCCCTCAGCTGGAAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTTATTCATCTCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTTATTCATCTCTAATGACCTC 900
Qy 901 CGACATCATGAGCCCTCAAGCCAAAGCTCTCTCTTCAAGGATAAGGAGCTAATTTGCCATCAAT 960
Db 901 CGACATCATGAGCCCTCAAGCCAAAGCTCTCTCTTCAAGGATAAGGAGCTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGAGTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGAGTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTACGCTGGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTACGCTGGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCTCCTCTGTGAAAGAAAGTACAGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCTCCTCTGTGAAAGAAAGTACAGGCTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
Qy 1261 ATGTA 1265
Db 1261 ATGCA 1265

RESULT 9
LOCUS AR653695
DEFINITION Sequence 9 from patent US 6887696.
ACCESSION AR653695

```

VERSION AR653695.1 GI:67584280
SOURCE .
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1296)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 9 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
    source
        1..1296
            /organism="unknown"
            /mol_type="genomic DNA"
ORIGIN
Query Match 99.8%; Score 1263.4; DB 6; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCCTGGCC 60
QY 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCCCT 120
DB 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCCCT 120
QY 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTTGACTGCCAGGAAGGCCA 180
QY 181 GATTCCTGCATCAGTCAGAGAGCTTTCATGGAGATGGCAGAGCTCATGCTCTCAGAGGC 240
DB 181 GATTCCTGCATCAGTCAGAGAGCTTTCATGGAGATGGCAGAGCTCATGCTCTCAGAGGC 240
QY 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGAGTGGCTCCCCAAGA 300
DB 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGAGTGGCTCCCCAAGA 300
QY 301 GATTCAGAGGCAAGATTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTGCCAGCTA 360
DB 301 GATTCAGAGGCAAGATTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTGCCAGCTA 360
QY 361 GCTAATTTGTTTCACAGCAAGAGACTGAGCTAGGATTTATGCAGATGTTGGAATAAA 420
DB 361 GCTAATTTGTTTCACAGCAAGAGACTGAGCTAGGATTTATGCAGATGTTGGAATAAA 420
QY 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATACCTAGCAATTCATGCCAGACCTTTGCT 480
DB 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATACCTAGCAATTCATGCCAGACCTTTGCT 480
QY 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTTGTTACTGTGACAGATTTGAAAAATTG 540
DB 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTTGTTACTGTGACAGATTTGAAAAATTG 540
QY 541 GCAGATGGTTATAGACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600
DB 541 GCAGATGGTTATAGACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTGGAAAAAGTAAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTGGAAAAAGTAAAG 720
QY 721 AGTATCTTGACCTGGACATCTTTTAAACCAGGAGAGAAATTTGATGTTGTCGACAGGG 780
DB 721 AGTATCTTGACCTGGACATCTTTTAAACCAGGAGAGAAATTTGATGTTGTCGACAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTGATTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
```

```

DB 781 GGTGGAAATGACCCAGATATGTTAGTGATTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTAATGACCTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCCTCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCCTCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGGTACAGCTTAGACAGGGAGACAATTTGAAGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGGTACAGCTTAGACAGGGAGACAATTTGAAGTGG 1020
QY 1021 GAACACCTCTCTCAGGCTTAGCGCTTGGCTGTAGCTATGATAAACCAGGAGATTGGT 1080
DB 1021 GAACACCTCTCTCAGGCTTAGCGCTTGGCTGTAGCTATGATAAACCAGGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
DB 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
QY 1141 GCCTGCTTCATCACAAGCTCTCCTGTGAAAAAGAAAGCTAGGTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCATCACAAGCTCTCCTGTGAAAAAGAAAGCTAGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTTCACATAATCCACAGCACTGTTTGTCTTCAGCTAGAAAAACACA 1260
DB 1201 TCAAGGTTAAGAGTTCACATAATCCACAGCACTGTTTGTCTTCAGCTAGAAAAACACA 1260
QY 1261 ATGTA 1265
DB 1261 ATGCA 1265
RESULT 10
LOCUS AR654963
DEFINITION Sequence 9 from patent US 6890748.
ACCESSION AR654963
VERSION AR654963.1 GI:67586502
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1296)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 9 10-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
    source
        1..1296
            /organism="unknown"
            /mol_type="genomic DNA"
ORIGIN
Query Match 99.8%; Score 1263.4; DB 6; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCCTGGCC 60
QY 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCCCT 120
DB 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCCCT 120
QY 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTTGACTGCCAGGAAGGCCA 180
QY 181 GATTCCTGCATCAGTCAGAGAGCTTTCATGGAGATGGCAGAGCTCATGCTCTCAGAGGC 240
DB 181 GATTCCTGCATCAGTCAGAGAGCTTTCATGGAGATGGCAGAGCTCATGCTCTCAGAGGC 240
QY 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGAGTGGCTCCCCAAGA 300
DB 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGAGTGGCTCCCCAAGA 300
QY 301 GATTCAGAGGCAAGATTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTGCCAGCTA 360
DB 301 GATTCAGAGGCAAGATTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTGCCAGCTA 360
QY 361 GCTAATTTGTTTCACAGCAAGAGACTGAGCTAGGATTTATGCAGATGTTGGAATAAA 420
DB 361 GCTAATTTGTTTCACAGCAAGAGACTGAGCTAGGATTTATGCAGATGTTGGAATAAA 420
QY 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATACCTAGCAATTCATGCCAGACCTTTGCT 480
DB 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATACCTAGCAATTCATGCCAGACCTTTGCT 480
QY 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTTGTTACTGTGACAGATTTGAAAAATTG 540
DB 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTTGTTACTGTGACAGATTTGAAAAATTG 540
QY 541 GCAGATGGTTATAGACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600
DB 541 GCAGATGGTTATAGACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTGGAAAAAGTAAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTGGAAAAAGTAAAG 720
QY 721 AGTATCTTGACCTGGACATCTTTTAAACCAGGAGAGAAATTTGATGTTGTCGACAGGG 780
DB 721 AGTATCTTGACCTGGACATCTTTTAAACCAGGAGAGAAATTTGATGTTGTCGACAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTGATTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
```


QY 661 CAGTACTGCAATCACCTGGCGAAATTTTGTGTCATTTGATGATTTCTCGGAAAGTATAAAG 720
Db |||||
QY 661 CAGTACTGCAATCACCTGGCGAAATTTTGTGTCATTTGATGATTTCTCGGAAAGTATAAAG 720
Db |||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db |||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db |||||
QY 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGAACTTTGGCTCAGCTGGATCAGCAA 840
Db |||||
QY 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGAACTTTGGCTCAGCTGGAAATCAGCAA 840
Db |||||
QY 841 GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTATTATGCTTAATGACCTC 900
Db |||||
QY 841 GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTATTATGCTTAATGACCTC 900
Db |||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAAATGCCATCAAT 960
Db |||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAAATGCCATCAAT 960
Db |||||
QY 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Db |||||
QY 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Db |||||
QY 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAAACCGGAGAGATTGGT 1080
Db |||||
QY 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAAACCGGAGAGATTGGT 1080
Db |||||
QY 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAAATCCT 1140
Db |||||
QY 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAAATCCT 1140
Db |||||
QY 1141 GCCTGCTTATACACAGCTCTCCCTGTGTAAGAGAACTAGGCTTCTATGAATGGACT 1200
Db |||||
QY 1141 GCCTGCTTATACACAGCTCTCCCTGTGTAAGAGAACTAGGCTTCTATGAATGGACT 1200
Db |||||
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260
Db |||||
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260
Db |||||
QY 1261 ATGTA 1265
Db |||||
QY 1261 ATGCA 1265
Db |||||

RESULT 12
AR654961
LOCUS AR654961 1308 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 5 from patent US 6890748.
ACCESSION AR654961
VERSION AR654961.1 GI:67586498
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1308)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 5 10-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source 1. .1308
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 99.8%; Score 1263.4; DB 6; Length 1308;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTCGCTTCCTGGCC 60
Db |||||
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTCGCTTCCTGGCC 60
Db |||||

QY 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACNAATGGATTCGCAAGGAGCGCT 120
Db |||||
QY 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTCGCAAGGAGCGCT 120
Db |||||
QY 121 ACCATGGGCTGGCTGCATCTGGGAGGCTTCATGTGCAACCTTTGACTGCGCAGGAAGAGCCA 180
Db |||||
QY 121 ACCATGGGCTGGCTGCATCTGGGAGGCTTCATGTGCAACCTTTGACTGCGCAGGAAGAGCCA 180
Db |||||
QY 181 GATTCTCTGATCAGTGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db |||||
QY 181 GATTCTCTGATCAGTGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db |||||
QY 241 TGGAGAGGATGCAAGGTTATGAGTACCTCTGCAATGATGATGCTGTGATGCTCCCAAGA 300
Db |||||
QY 241 TGGAGAGGATGCAAGGTTATGAGTACCTCTGCAATGATGATGCTGTGATGCTCCCAAGA 300
Db |||||
QY 301 GATTCTCAGAGGAGCAGACTTTCAGGAGAGCCCTCAGCGCTTTCTTCATGGGATTCGCCAGCTA 360
Db |||||
QY 301 GATTCTCAGAGGAGCAGACTTTCAGGAGAGCCCTCAGCGCTTTCTTCATGGGATTCGCCAGCTA 360
Db |||||
QY 361 GCTAATTTATGTTCA CAGCAAGGACTGAAGCTAGGAGTTTATGCAAGATGTTGGAAATAAA 420
Db |||||
QY 361 GCTAATTTATGTTCA CAGCAAGGACTGAAGCTAGGAGTTTATGCAAGATGTTGGAAATAAA 420
Db |||||
QY 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACAGCACTTGTGATGCTGAGAGTTTGGAAATTTG 480
Db |||||
QY 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACAGCACTTGTGATGCTGAGAGTTTGGAAATTTG 480
Db |||||
QY 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGTTGTTACTGCTGACAGTTTGGAAATTTG 540
Db |||||
QY 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGTTGTTACTGCTGACAGTTTGGAAATTTG 540
Db |||||
QY 541 CGAGATGTTTAAAGACATGCTTGGCCCTGGAATAGGACTGGCAGAGCAATTTGTGTAC 600
Db |||||
QY 541 CGAGATGTTTAAAGACATGCTTGGCCCTGGAATAGGACTGGCAGAGCAATTTGTGTAC 600
Db |||||
QY 601 TCCTGCTGAGTGGCTCTTTATATGTTGGCTTTTAAAGCCCAATTTATACAGAAATCCGA 660
Db |||||
QY 601 TCCTGCTGAGTGGCTCTTTATATGTTGGCTTTTAAAGCCCAATTTATACAGAAATCCGA 660
Db |||||
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATTTCTCGGAAAGTATAAAG 720
Db |||||
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATTTCTCGGAAAGTATAAAG 720
Db |||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGGACAGGG 780
Db |||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGGACAGGG 780
Db |||||
QY 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCTCAGCTGGAAATCAGCAA 840
Db |||||
QY 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCTCAGCTGGAAATCAGCAA 840
Db |||||
QY 841 GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTATTATGCTTAATGACCTC 900
Db |||||
QY 841 GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTATTATGCTTAATGACCTC 900
Db |||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAAATGCCATCAAT 960
Db |||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAAATGCCATCAAT 960
Db |||||
QY 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Db |||||
QY 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Db |||||
QY 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAAACCGGAGAGATTGGT 1080
Db |||||
QY 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAAACCGGAGAGATTGGT 1080
Db |||||
QY 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAAATCCT 1140
Db |||||
QY 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAAATCCT 1140
Db |||||

```
QY 1141 GCCTGCTTCATCACAGACTCCTCCCTGTGAAAGAGAGCTAGGTTCTATGAATGGACT 1200
Db |||||
QY 1141 GCCTGCTTCATCACAGACTCCTCCCTGTGAAAGAGAGCTAGGTTCTATGAATGGACT 1200
Db |||||
QY 1201 TCAAGGTTAAGAGTACACATAATCCACAGGACCTGTTTCTTCAGCTAGAAAACACA 1260
Db |||||
QY 1201 TCAAGGTTAAGAGTACACATAATCCACAGGACCTGTTTCTTCAGCTAGAAAACACA 1260
Db |||||
QY 1261 ATGTA 1265
Db |||||
QY 1261 ATGCA 1265
Db |||||

RESULT 13
CS135670
LOCUS Homo sapiens
DEFINITION Homo sapiens
ACCESSION CS135670
VERSION CS135670.1 GI:72056330
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Garman,S.C.
AUTHORS Crystal structure of human a-galactosidase
TITLE Patent: WO 2005069192-A 1 28-JUL-2005;
JOURNAL Transkaryotic Therapies, Inc. (US); Selden, Richard F. (US);
Garbocci, David N. (US); Treco, Douglas A. (US); NIH/NIAD (US);
Garman, Scott C. (US); Borowski, Marianne (US); Kinoshita, Carol M.
(US)
FEATURES
source Location/Qualifiers
1..1290
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 99.7%; Score 1261.8; DB 6; Length 1290;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1263; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCAAGCTGAGAAACCCAGAACTACATCTGGGCTGGGCTTGGCTTGGCTTCTCTGGCC 60
Db |||||
QY 1 ATGCAAGCTGAGAAACCCAGAACTACATCTGGGCTGGGCTTGGCTTGGCTTCTCTGGCC 60
Db |||||
QY 61 CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
Db |||||
QY 61 CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
Db |||||
QY 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180
Db |||||
QY 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180
Db |||||
QY 181 GATTCCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGCTTCAGAGGC 240
Db |||||
QY 181 GATTCCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGCTTCAGAGGC 240
Db |||||
QY 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db |||||
QY 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db |||||
QY 301 GATTCAGAGGAGAGCTTCAGGACAGCCCTCAGCCCTTCTCATGGATGGCTCCAGGCTA 360
Db |||||
QY 301 GATTCAGAGGAGAGCTTCAGGACAGCCCTCAGCCCTTCTCATGGATGGCTCCAGGCTA 360
Db |||||
QY 361 GCTAATATGTTTACAGCAAGAGCACTGAGAGCTAGGATTTATGACAGATGTTGGAAATAA 420
Db |||||
QY 361 GCTAATATGTTTACAGCAAGAGCACTGAGAGCTAGGATTTATGACAGATGTTGGAAATAA 420
Db |||||
QY 421 ACCTGGCAGGCTTCCCTGGAGTTTGGATACTAGACATTTGATGCCAGACCTTTGCT 480
Db |||||
```

```
Db |||||
QY 421 ACCTGGCAGGCTTCCCTGGAGTTTGGATACTAGCATTTGATGCCAGACCTTTGCT 480
Db |||||
QY 481 GACTGGGAGTAGATCTCTAAATTTTCTGATGTTTCTGATGAGTGTGAGATTTGAAAATTG 540
Db |||||
QY 481 GACTGGGAGTAGATCTCTAAATTTTCTGATGTTTCTGATGAGTGTGAGATTTGAAAATTG 540
Db |||||
QY 541 CGAGATGTTTATAAGCACATGCTCTGGCCCTGTAATAGGACTGGCAGAGCAATTTGTAC 600
Db |||||
QY 541 CGAGATGTTTATAAGCACATGCTCTGGCCCTGTAATAGGACTGGCAGAGCAATTTGTAC 600
Db |||||
QY 601 TCCTGTGAGTGGCTCTTTTATATGTTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db |||||
QY 601 TCCTGTGAGTGGCTCTTTTATATGTTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db |||||
QY 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTGGAAGATTAAG 720
Db |||||
QY 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTGGAAGATTAAG 720
Db |||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db |||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db |||||
QY 781 GGTGGATGACCCAGATATGTTAGTGTGCACTTTGGCCTCAGCTGGAAATCAGCAA 840
Db |||||
QY 781 GGTGGATGACCCAGATATGTTAGTGTGCACTTTGGCCTCAGCTGGAAATCAGCAA 840
Db |||||
QY 841 GTAACCTGAGTGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTAATCAGCTC 900
Db |||||
QY 841 GTAACCTGAGTGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTAATCAGCTC 900
Db |||||
QY 901 CGACATCAAGCTCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db |||||
QY 901 CGACATCAAGCTCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db |||||
QY 961 CAGACCCCTTGGGCAAGAGGTACAGCTTAGACAGGAGAGCAACTTTGAAGTGG 1020
Db |||||
QY 961 CAGACCCCTTGGGCAAGAGGTACAGCTTAGACAGGAGAGCAACTTTGAAGTGG 1020
Db |||||
QY 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAGCCGAGGAGATGGT 1080
Db |||||
QY 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAGCCGAGGAGATGGT 1080
Db |||||
QY 1081 GGACCTCGCTCTTATACCATCGAGTGTCTTCCCTGGGTAAGGAGTGGCTGTAACTCT 1140
Db |||||
QY 1081 GGACCTCGCTCTTATACCATCGAGTGTCTTCCCTGGGTAAGGAGTGGCTGTAACTCT 1140
Db |||||
QY 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGAGAGCTAGGTTCTATGAATGGACT 1200
Db |||||
QY 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGAGAGCTAGGTTCTATGAATGGACT 1200
Db |||||
QY 1201 TCAAGGTTAAGAGTACACATAATCCACAGGACCTGTTTCTTCAGCTAGAAAACACA 1260
Db |||||
QY 1201 TCAAGGTTAAGAGTACACATAATCCACAGGACCTGTTTCTTCAGCTAGAAAACACA 1260
Db |||||
QY 1261 ATGTA 1265
Db |||||
QY 1261 ATGCA 1265
Db |||||

RESULT 14
BT006864
LOCUS Homo sapiens galactosidase, alpha mRNA linear PRI 13-MAY-2003
DEFINITION Homo sapiens galactosidase, alpha mRNA, complete cds.
ACCESSION BT006864
VERSION BT006864.1 GI:30582566
KEYWORDS FLI CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1290)
```



```

ACCESSION AY898631
VERSION AY898631.1 GI:60655716
KEYWORDS Human ORF Project.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 1290)
AUTHORS Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
Williamson,J. and LaBaer,J.
TITLE Cloning of human full-length CDS in Creator (TM) recombinational
vector system
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1290)
AUTHORS Hines,L., Rolfs,A., Jepson,D., Moreira,D., Raphael,J., Kelley,F.,
Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
Williamson,J. and LaBaer,J.
DIRECT SUBMISSION
TITLE Submitted (04-JAN-2005) Biological Chemistry and Molecular
JOURNAL Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
Cambridge, MA 02141, USA
COMMENT This CDS clone is a part of a collection of human full-length
expression clones generated by Harvard Institute of Proteomics.
This ORF clone has been cloned with normalized stop-codon. The CDS
has been directionally cloned using BD In-Fusion(TM) cloning system
between the SalI and HindIII sites of the pDNR-Dual vector.
Additional sequences in the clone: 'ACC' after SalI site and
before 'ATG' to provide Kozak consensus sequence. Each clone is
clonally isolated and full-length sequence-verified.
FEATURES
    Location/Qualifiers
        1..1290
            /organism="synthetic construct"
            /mol_type="mRNA"
            /db_xref="taxon:32630"
            /clone="FLH025769.01X"
            /lab_host="Escherichia coli DH5alpha Tl resistant"
            /note="derived from MGC template"
        1..1290
            /gene="GLA"
            /gene="GLA"
            /codon_start=1
            /transl_table=1
            /product="galactosidase alpha"
            /protein_id="AA32422.1"
            /db_xref="GI:60655717"
            /translation="MQLRNPGLGCLALRLALVSWDIPGARALDNLGARTPTMGW
            LWERFMCNLDQEBPDSCISEKLPMENALMVSEGWKADGVEYLICIDDCWNPORDS
            EGRQADQRFPHGIRQLANYVHSGKLIGIYADVGNKTCAGFPGSFGYDIDDAFA
            DWGVLLKPDGVCDSLENLADGYKMSLNRGTSIVYSCWELYPMPFKPNYTE
            IRQYCNHWNFADIDDSWKSISILDWTSFNQERIVDVAGPGWMDPMLVGNFGLS
            WNQOVTOMALWAIMAAPLPMNDLRHISPOAKALLQDKDVIAINODPLGKGYQLRQG
            DNFEWEPPLSLGLAWAMINROEIGGPRSTIIVASLKGWACNPACFIQLLPVKR
            KLGFVWTSRLRSHINPTGTVLQLLENTQMQLKOLL"
ORIGIN
Query Match 99.7%; Score 1261.8; DB 11; Length 1290;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1263; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
QY 61 CTGCTTCTGGGACATCCTGGGCTAGACACTGGACAATGGATTGGCAGAGCGCT 120
DB 61 CTGCTTCTGGGACATCCTGGGCTAGACACTGGACAATGGATTGGCAGAGCGCT 120
QY 121 ACCATGGCTGGCTGGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGCTGGCTGGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCCA 180
QY 181 GATTCCTGCATCAGTGAGAAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

```

Db 1261 ATGCA 1265

Search completed: December 26, 2005, 07:09:28
Job time : 6371 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 01:02:18 ; Search time 771 Seconds
(without alignments)
10943.581 Million cell updates/sec

US-10-602-219-11

Title: Perfect score: 1266
Sequence: 1 atgcagctgaggaaccacaga.....agctagaacacacaaatgtaa 1266

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	100.0	1266	6	AAD45223
2	1266	100.0	1266	10	ADD84750 Human rGA
3	1266	100.0	1266	12	ADJ88276 Human WT
4	1266	100.0	1266	12	ADM48680 Human wil
5	1266	100.0	1266	13	ADU66915 Human alp
6	1266	100.0	1266	14	AEA27444 Human alp
7	1266	99.8	1284	6	AAD45224
8	1264	99.8	1284	10	ADD84752 Human rGA
9	1264	99.8	1284	12	ADJ88278 Human WT
10	1264	99.8	1284	12	ADM48682 Human wil
11	1264	99.8	1284	13	ADU66917 Human alp
12	1264	99.8	1284	14	AEA27446 Human alp
13	1263.4	99.8	1278	6	AAD45221
14	1263.4	99.8	1278	10	ADD84746 Human alp
15	1263.4	99.8	1278	12	ADJ88272 Human rGA
16	1263.4	99.8	1278	12	ADM48676 Human wil
17	1263.4	99.8	1278	13	ADU66911 Human alp
18	1263.4	99.8	1278	14	AEA27440 Human alp
19	1263.4	99.8	1290	6	AAD45219 Human WT

20	1263.4	99.8	1290	10	ADD84742	Add84742 Human alp
21	1263.4	99.8	1290	12	ADJ88268	Adj88268 Human WT
22	1263.4	99.8	1290	12	ADM48672	Adm48672 Human wil
23	1263.4	99.8	1290	13	ADU66907	Adu66907 Human alp
24	1263.4	99.8	1290	14	AEA27436	Aea27436 Human alp
25	1263.4	99.8	1296	6	AAD45222	Aad45222 Human rGA
26	1263.4	99.8	1296	10	ADD84748	Add84748 Human alp
27	1263.4	99.8	1296	12	ADJ88274	Adj88274 Human WT
28	1263.4	99.8	1296	12	ADM48678	Adm48678 Human wil
29	1263.4	99.8	1296	13	ADU66913	Adu66913 Human alp
30	1263.4	99.8	1296	14	AEA27442	Aea27442 Human alp
31	1263.4	99.8	1308	6	AAD45220	Aad45220 Human WT-
32	1263.4	99.8	1308	10	ADD84744	Add84744 Human alp
33	1263.4	99.8	1308	12	ADJ88270	Adj88270 Human WT
34	1263.4	99.8	1308	12	ADM48674	Adm48674 Human wil
35	1263.4	99.8	1308	13	ADU66909	Adu66909 Human alp
36	1263.4	99.8	1308	14	AEA27438	Aea27438 Human alp
37	1261.8	99.7	1290	12	ADN49739	Adn49739 Human alp
38	1261.8	99.7	1290	13	ADU74415	Adu74415 Human alp
39	1261.8	99.7	1290	14	AEA3187	Aeb43187 Human alp
40	1261.8	99.7	1306	8	ABZ79777	Abz79777 Human alp
41	1261.8	99.7	1319	10	ACF80581	Acf80581 Human alp
42	1261.8	99.7	1339	10	ADH54464	Adh54464 Human alp
43	1261.8	99.7	1343	2	AAV31801	Aav31801 Nucleotid
44	1261.8	99.7	1343	3	AAA75434	Aaa75434 Nucleotid
45	1261.8	99.7	1343	3	AAA70685	Aaa70685 Human alp

ALIGNMENTS

RESULT 1

AAD45223

ID AAD45223 standard; DNA; 1266 BP.

XX AAD45223;

DT 27-DEC-2002 (first entry)

XX Human rGAL-8 DNA.

XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-8; Gene; ds.

XX Homo sapiens.

OS Location/Qualifiers

FH Key 1..1266

FT CDS /tag= a

FT /product= "Human rGAL-8 protein"

XX US2002088024-A1.

XX 04-JUL-2002.

XX 13-NOV-2001; 2001US-00993059.

XX 26-JUL-2000; 2000US-00626127.

XX (GARG/) GARGER S J.

XX (TURP/) TURPEN T H.

XX (KUNAI/) KUMAGAI M H.

XX Garger SJ, Turpen TH, Kumagai MH;

XX WPI; 2002-681656/73.

XX P-PSDB; AAE28210.

XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal storage diseases.

XX Claim 1; Page 42-44; 88pp; English.

XX

PT lysosomal enzymes in a plant expression system that can be used in enzyme replacement therapy.

XX Claim 1; SEQ ID NO 11; 77pp; English.

XX The invention relates to human alpha-galactosidase derivatives and the CC nucleic acids encoding them. The polypeptides are used in a method for CC producing active recombinant human and animal lysosomal enzymes in a CC plant expression system. The enzymes can be used in enzyme replacement CC therapy for the therapeutic treatment of human and animal lysosomal CC diseases. This sequence represents DNA encoding a human alpha-galactosidase derivative polypeptide of the invention.

XX SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Query Match 100.0%; Score 1266; DB 10; Length 1266;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCACTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
 DB 1 ATGCACTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
 QY 61 CTCGTTTCTGGGACATCTCTGGGCTAGAGCACTGGCAATGGATTGGCAGGCGCT 120
 DB 61 CTCGTTTCTGGGACATCTCTGGGCTAGAGCACTGGCAATGGATTGGCAGGCGCT 120
 QY 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTGCCTCCAGGAGGCCA 180
 DB 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTGCCTCCAGGAGGCCA 180
 QY 181 GATTCCTGTCATCAGTGAGAGCTCTTTCATGGAGTGGCAGAGCTCATGTCTCAGAAGGC 240
 DB 181 GATTCCTGTCATCAGTGAGAGCTCTTTCATGGAGTGGCAGAGCTCATGTCTCAGAAGGC 240
 QY 241 TGGAGGATGCAGTTATGAGTACCTCTGCTGATGATGACTGTTGGATGCTCCCAAGA 300
 DB 241 TGGAGGATGCAGTTATGAGTACCTCTGCTGATGATGACTGTTGGATGCTCCCAAGA 300
 QY 301 GATTCAGAGGAGACTTCAGGAGAGCTTCAGGAGAGCTTCCTCATGGGATTCGCCAGCTA 360
 DB 301 GATTCAGAGGAGACTTCAGGAGAGCTTCAGGAGAGCTTCCTCATGGGATTCGCCAGCTA 360
 QY 361 GCTAATATATGTTTACAGCAAGAGCTGAAGCTAGGATTTATGAGATGTTGGAATAAA 420
 DB 361 GCTAATATATGTTTACAGCAAGAGCTGAAGCTAGGATTTATGAGATGTTGGAATAAA 420
 QY 421 ACTTGGCAGGCTTCCCTGGGAGTTTGGATACGACATTCGATTCGCCAGACCTTTGCT 480
 DB 421 ACTTGGCAGGCTTCCCTGGGAGTTTGGATACGACATTCGATTCGCCAGACCTTTGCT 480
 QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
 DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
 QY 541 GCAGATGGTTATAGACACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTATC 600
 DB 541 GCAGATGGTTATAGACACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTATC 600
 QY 601 TCTGTGAGTGGCTCTTTATATGTTGGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
 DB 601 TCTGTGAGTGGCTCTTTATATGTTGGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
 QY 661 CAGTACTGCAATCAGTGGGAAATTTTGTGACATTTGATGATTTTCCCTGGAATAAG 720
 DB 661 CAGTACTGCAATCAGTGGGAAATTTTGTGACATTTGATGATTTTCCCTGGAATAAG 720
 QY 721 AGTATCTTGACCTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTTGCTGGACAGGG 780
 DB 721 AGTATCTTGACCTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTTGCTGGACAGGG 780
 QY 781 GGTGGAATGACCCAGATATGTTAGTATGGCACTTTGGCTTCAGCTGGAATCAGCAA 840
 DB 781 GGTGGAATGACCCAGATATGTTAGTATGGCACTTTGGCTTCAGCTGGAATCAGCAA 840

Db 781 GGTGGAATGACCCAGATATGTTAGTATGGCACTTTGGCTTCAGCTGGAATCAGCAA 840
 QY 841 GTAACCTAGATGGCCCTCTGGGCTATCATGGCTATCATGGCTCTCTTTATCATGTCTAATGACCTC 900
 Db 841 GTAACCTAGATGGCCCTCTGGGCTATCATGGCTATCATGGCTCTCTTTATCATGTCTAATGACCTC 900
 QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGACGTAATTTGCCATCAAT 960
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGACGTAATTTGCCATCAAT 960
 QY 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTTAGACAGGGAGACAACTTTCAAGTGTGG 1020
 Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTTAGACAGGGAGACAACTTTCAAGTGTGG 1020
 QY 1021 GAAACGACCTCTCTCAGGCTTACGCTGGGCTGTAGCTATGATATAAACCGGACGAGATTGGT 1080
 Db 1021 GAAACGACCTCTCTCAGGCTTACGCTGGGCTGTAGCTATGATATAAACCGGACGAGATTGGT 1080
 QY 1081 GGACCTCGCTCTTATACCATGCGAGTTGCTTCCCTGGGTAAAGGATGCGCTGTATCTCT 1140
 Db 1081 GGACCTCGCTCTTATACCATGCGAGTTGCTTCCCTGGGTAAAGGATGCGCTGTATCTCT 1140
 QY 1141 GCTCTCTTCATCACACAGCTCTCTCCCTGTGAAAAGGAAAGCTAGGTTCTATGAAATGGACT 1200
 Db 1141 GCTCTCTTCATCACACAGCTCTCTCCCTGTGAAAAGGAAAGCTAGGTTCTATGAAATGGACT 1200
 QY 1201 TCAAGGTTAAGAGTACATTAATCCACAGGCTGTTTTCCTTTCAGCTAGAAAACACA 1260
 Db 1201 TCAAGGTTAAGAGTACATTAATCCACAGGCTGTTTTCCTTTCAGCTAGAAAACACA 1260
 QY 1261 ATGTAA 1266
 Db 1261 ATGTAA 1266
 RESULT 3
 ADJ88276
 ID ADJ88276 standard; DNA; 1266 BP.
 XX AC ADJ88276;
 XX DT 06-MAY-2004 (first entry)
 XX DE Human WT rGAL-8 (galactosidase) DNA.
 XX KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
 KW Fabry's disease; Gaucher's disease; human; gene; ds.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 XX CDS 1..1266
 XX FT /*tag= a
 XX FT /product= "Human WT rGAL-8 protein"
 XX US2004016021-A1.
 XX PN 22-JAN-2004.
 XX PD 23-JUN-2003; 2003US-00602219.
 XX PP 26-FEB-1988; 88US-00160766.
 XX PR 26-FEB-1988; 88US-00160771.
 XX PR 15-JUL-1988; 88US-00219279.
 XX PR 17-FEB-1989; 89US-00310881.
 XX PR 05-MAY-1989; 89US-00347637.
 XX PR 08-JUN-1989; 89US-00363138.
 XX PR 22-OCT-1990; 90US-00600244.
 XX PR 16-JAN-1991; 91US-00641617.
 XX PR 26-JUL-1991; 91US-00737899.
 XX PR 01-AUG-1991; 91US-00739143.
 XX PR 31-JUL-1992; 92US-00923692.
 XX PR 30-DEC-1992; 92US-00997733.

PR 29-DEC-1993; 93US-00176414.
 PR 19-JAN-1994; 94US-00184237.
 PR 14-OCT-1994; 94US-00324003.
 PR 21-MAY-1999; 99US-00316572.
 PR 26-JUL-2000; 2000US-00626127.
 PR 13-NOV-2001; 2001US-00993059.
 XX
 PA (TURP/) TURPEN T H.
 PA (POGU/) POGUE G P.
 PA (ERWI/) ERWIN R L.
 PA (GRIL/) GRILL L K.
 XX
 PI Turpen TH, Pogue GP, Erwin RL, Grill LK;
 XX
 XX WPI; 2004-108227/11.
 DR P-PSDB; ADJ88277.
 XX
 PT New lysosomal enzymes, useful in treating human and animal lysosomal
 PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
 XX
 XX Claim 1; SEQ ID NO 11; 71pp; English.
 XX
 CC The invention relates to nucleotide encoding galactosidase (GAL). The
 CC invention is useful in gene therapy. The polynucleotides and polypeptides
 CC are useful in treating human and animal lysosomal storage diseases, e.g.
 CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
 CC DNA.
 XX
 SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
 Query Match 100.0%; Score 1266; DB 12; Length 1266;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTCCTGGCC 60
 Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTCCTGGCC 60
 QY 61 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGGATGGCAAGAGCGCT 120
 Db 61 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGGATGGCAAGAGCGCT 120
 QY 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTCACCTGCCAGGAAGGCCA 180
 Db 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTCACCTGCCAGGAAGGCCA 180
 QY 181 GATTCCTGCATCAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
 Db 181 GATTCCTGCATCAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
 QY 241 TGAAGGATGCAGGTATAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAAGA 300
 Db 241 TGAAGGATGCAGGTATAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAAGA 300
 QY 301 GATTCAGAAGGACAGCTTCAGGAGAGCCCTCAGCGCTTTCTCATGGATGGCCAGCTA 360
 Db 301 GATTCAGAAGGACAGCTTCAGGAGAGCCCTCAGCGCTTTCTCATGGATGGCCAGCTA 360
 QY 361 GCTAAATATGTTCCACAGCAAGGAGCTGAAGCTAGGAGTTATGCAGATGTTGAAATATAA 420
 Db 361 GCTAAATATGTTCCACAGCAAGGAGCTGAAGCTAGGAGTTATGCAGATGTTGAAATATAA 420
 QY 421 ACCTGGCAGGCTTCCTGGAGTTTGGATACACAGATTCAGATCCAGCCAGCTTTGCT 480
 Db 421 ACCTGGCAGGCTTCCTGGAGTTTGGATACACAGATTCAGATCCAGCCAGCTTTGCT 480
 QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTACACAGATTTGGAAATTTG 540
 Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTACACAGATTTGGAAATTTG 540
 QY 541 GCAGATGGTTATAGCAATGTCTCTGGCCCTGAATAGGACTGGCAGAGAGCATTTGTATC 600
 Db 541 GCAGATGGTTATAGCAATGTCTCTGGCCCTGAATAGGACTGGCAGAGAGCATTTGTATC 600

QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCCA 660
 Db 601 TCCTGTGAGTGGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCCA 660
 QY 661 CAGTACTGCAATCACTGCGGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATAAAG 720
 Db 661 CAGTACTGCAATCACTGCGGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATAAAG 720
 QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTCCTGGACAGGG 780
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTCCTGGACAGGG 780
 QY 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCCCTCAGCTGGGAATCAGCAA 840
 Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCCCTCAGCTGGGAATCAGCAA 840
 QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTAATGACCTC 900
 Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTAATGACCTC 900
 QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
 Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
 QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
 Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
 QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCGGAGGAGATTGGT 1080
 Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCGGAGGAGATTGGT 1080
 QY 1081 GGACCTGCTCTTATACCATCGAGTGTCTCCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
 Db 1081 GGACCTGCTCTTATACCATCGAGTGTCTCCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
 QY 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAGTGGGTTCTATGAATGGACT 1200
 Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAGTGGGTTCTATGAATGGACT 1200
 QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGACCTGTTTTGCTTCAGCTAGAAAACACA 1260
 Db 1201 TCAAGGTTAAGAGTACATAAATCCACAGGACCTGTTTTGCTTCAGCTAGAAAACACA 1260
 QY 1261 ATGTAA 1266
 Db 1261 ATGTAA 1266
 RESULT 4
 ADM48680
 ID ADM48680 standard; DNA; 1266 BP.
 XX
 AC ADM48680;
 XX
 DT 03-JUN-2004 (first entry)
 XX
 DE Human wild type rGAL-8 DNA.
 XX
 KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
 KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
 KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
 KW gene; db.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1266
 FT /*tag= a
 FT /product= "GAL-8 protein"
 XX
 PN US2004023281-A1.

[illegible]


```
Db      1201  TCAAGGTAAAGTACATATAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
QY      1261  ATGTAA 1266
Db      1261  ATGTAA 1266

RESULT 7
AAD45224
ID      AAD45224 standard; DNA; 1284 BP.
AC      AAD45224;
XX      27-DEC-2002 (first entry)
DT      Human rGAL-8R DNA.
DE      Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW      therapeutic; rGAL-8R; gene; ds.
XX      Homo sapiens.
XX      Key
FH      Location/Qualifiers
FT      1. .1284
FT      /*tag= a
FT      /product= "Human rGAL-8R protein"
XX      US2002088024-A1.
XX      04-JUL-2002.
XX      13-NOV-2001; 2001US-00930359.
XX      26-JUL-2000; 2000US-00626127.
XX      (GARG/) GARGER S J.
XX      (TURP/) TURPEN T H.
XX      (KUMA/) KUMAGAI M H.
XX      Garger SJ, Turpen TH, Kumagai MH;
XX      WPI; 2002-681656/73.
XX      P-PSDB; AAE28211.
XX      Novel human alpha-galactosidase polypeptide useful for treating lysosomal
XX      storage diseases.
XX      Claim 1; Page 44-46; 88pp; English.
XX      The invention relates to human alpha-galactosidase truncated at the
XX      carboxy terminus and the production of enzymatically active recombinant
XX      human and animal lysosomal enzymes. The invention is useful for producing
XX      human and animal lysosomal enzymes for treating lysosomal storage diseases, producing
XX      altered or mutated proteins, enzymatically active or otherwise, to serve
XX      as precursors or substrates for further in vivo or in vitro processing to
XX      a specialised industrial form for research or therapeutic uses, to
XX      produce more effective therapeutic enzyme, for producing antibodies
XX      against lysosomal enzymes for medical diagnostic use, and in any
XX      commercial process that involves substrate hydrolysis. The present
XX      sequence is human rGAL-8R DNA
XX      Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;
XX      Query Match 99.8%; Score 1264; DB 6; Length 1284;
XX      Best Local Similarity 100.0%; Pred. No. 0;
XX      Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCGCTTCGCTTCCTGGCC 60
Db      1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCGCTTCGCTTCCTGGCC 60
QY      61 CTCGTTTCTCGGGACATCCCTGGGCTAGAGCACTGGCAATGATGGATGGCAAGGCGCCT 120
```

```
Db      61 CTCGTTTCTCGGGACATCCCTGGGCTAGAGCACTGGCAATGATGGATGGCAAGGCGCCT 120
QY      121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTCAACCTTTGACTGCCAGGAAGAGCCA 180
Db      121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTCAACCTTTGACTGCCAGGAAGAGCCA 180
QY      181 GATTCTCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTTCAGAAGGC 240
Db      181 GATTCTCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTTCAGAAGGC 240
QY      241 TCGAAGGATGCAGGTTATGAGTACCTCTGCATTTGATGACTGTTGATGGCTCCCAAGA 300
Db      241 TCGAAGGATGCAGGTTATGAGTACCTCTGCATTTGATGACTGTTGATGGCTCCCAAGA 300
QY      301 GATTTCAGAAGSCAGACTTCAGGCAGACCTTCAGCCCTTCAGCGCTTCCTCATGGGATTCGCCAGCTA 360
Db      301 GATTTCAGAAGSCAGACTTCAGGCAGACCTTCAGCCCTTCAGCGCTTCCTCATGGGATTCGCCAGCTA 360
QY      361 GCTAATTATGTTTCACAGCAAAAGGACTGAAGCTAGGGAATTTATGCAGATGTTGGAATAAAA 420
Db      361 GCTAATTATGTTTCACAGCAAAAGGACTGAAGCTAGGGAATTTATGCAGATGTTGGAATAAAA 420
QY      421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACCTACGACATTTGATGATGGCTCCAGACCTTTTGGT 480
Db      421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACCTACGACATTTGATGATGGCTCCAGACCTTTTGGT 480
QY      481 GACTGGGAGTAGATCTGCTAAATTTGATGCTGTTACTGTGACAGATTTTGGAAATTTG 540
Db      481 GACTGGGAGTAGATCTGCTAAATTTGATGCTGTTACTGTGACAGATTTTGGAAATTTG 540
QY      541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Db      541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
QY      601 TCCTGTGAGTGGCCTCTTTATATGATGGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db      601 TCCTGTGAGTGGCCTCTTTATATGATGGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY      661 CAGTACTGCAATCAGCTGGCGAAATTTTGTGCTGACATTTGATGATTCCTGGAAAGATATAAAG 720
Db      661 CAGTACTGCAATCAGCTGGCGAAATTTTGTGCTGACATTTGATGATTCCTGGAAAGATATAAAG 720
QY      721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTTGCTGACAGAGG 780
Db      721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTTGCTGACAGAGG 780
QY      781 GGTGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Db      781 GGTGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
QY      841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTTATTCTATGCTATGACCTC 900
Db      841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTTATTCTATGCTATGACCTC 900
QY      901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db      901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY      961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db      961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY      1021 GRACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGCAGGAGATTTGGT 1080
Db      1021 GRACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGCAGGAGATTTGGT 1080
QY      1081 GGACCTCGCTCTTTATACCATCGCAGTTGCTTCCTGGGTAAGAGAGTGGCCTGTAATCCT 1140
Db      1081 GGACCTCGCTCTTTATACCATCGCAGTTGCTTCCTGGGTAAGAGAGTGGCCTGTAATCCT 1140
QY      1141 GCTGCTTCATACACAGCTCCTCCTGTGTAAAGAGCTAGGCTTCTATGATGGACT 1200
```



```
Db 1141 GCCTGCTTCATCACAGACTCCTCCCTGTGAAAGAGCTAGGTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAACACA 1260
Qy 1261 ATGT 1264
Db 1261 ATGT 1264

RESULT 8
ADD84752
ID ADD84752 standard; DNA; 1284 BP.
XX
AC ADD84752;
XX
DE 29-JAN-2004 (first entry)
XX
DE Human alpha-galactosidase rGAL-8R DNA.
XX
KW Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1284
FT FT /*tag= a
FT FT /product= "Human rGAL-8R"
XX
PN US2003106095-A1.
XX
XX 05-JUN-2003.
XX
XX 20-MAR-2002; 2002US-00103327.
XX
XX 26-JUL-2000; 2000US-00626127.
XX
XX 13-NOV-2001; 2001US-00993059.
XX
XX (GARG/) GARGER S J.
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI: 2003-801257/75.
XX P-PSDB; ADD84753.
XX
XX New polynucleotide for producing active recombinant human and animal
XX lysosomal enzymes in a plant expression system that can be used in enzyme
XX replacement therapy.
XX
XX Claim 1; SEQ ID NO 13; 77pp; English.
XX
XX The invention relates to human alpha-galactosidase derivatives and the
XX nucleic acids encoding them. The polypeptides are used in a method for
XX producing active recombinant human and animal lysosomal enzymes in a
XX plant expression system. The enzymes can be used in enzyme replacement
XX therapy for the therapeutic treatment of human and animal lysosomal
XX diseases. This sequence represents DNA encoding a human alpha-
XX galactosidase derivative polypeptide of the invention.
XX
XX Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;
XX

Query Match 99.8%; Score 1284; DB 10; Length 1284;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
```

```
Qy 61 CTCGTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTTGGCAAGGAGCGCT 120
Db 61 CTCGTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTTGGCAAGGAGCGCT 120
Qy 121 ACCATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
Db 121 ACCATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
Qy 181 GATTCCTGTCATCAGTGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCCTGTCATCAGTGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATGATGATGATGATGATGATGATGATGATGAT 300
Db 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATGATGATGATGATGATGATGATGATGATGAT 300
Qy 301 GATTCAGAGGAGCAGACTTTCAGGAGAGCTTTCAGGAGAGCTTTCAGGAGAGCTTTCAGGAG 360
Db 301 GATTCAGAGGAGCAGACTTTCAGGAGAGCTTTCAGGAGAGCTTTCAGGAGAGCTTTCAGGAG 360
Qy 361 GCTAATTATGTTTACAGCAAGGAGCTGAAAGCTAGGAGTTTATGCGAGATGTTGGAAATAAA 420
Db 361 GCTAATTATGTTTACAGCAAGGAGCTGAAAGCTAGGAGTTTATGCGAGATGTTGGAAATAAA 420
Qy 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATGATGATGATGATGATGATGATGATGATGATGAT 480
Db 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATGATGATGATGATGATGATGATGATGATGATGAT 480
Qy 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
Qy 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGAGCAATTTGTGTAC 600
Db 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGAGCAATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTTATATGATGAGGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTTATATGATGAGGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATTTGATGATTTCTTGGGAAAGTATAAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATTTGATGATTTCTTGGGAAAGTATAAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGAAATTTGTGTGTTCTTGGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGAAATTTGTGTGTTCTTGGACAGGG 780
Qy 781 GGTGGAATGACCCAGATATGTTAGTATGTCATCTTTGGCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTATGTCATCTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTTATTATGATGCTATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTTATTATGATGCTATGACCTC 900
Qy 901 GCACACATCAGCCCTCAAGCCCAAGAGCTCTCTTTCAGGATAAGAGCTTAATTCGATCAAT 960
Db 901 GCACACATCAGCCCTCAAGCCCAAGAGCTCTCTTTCAGGATAAGAGCTTAATTCGATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAAACACCTCTCAGGCTTAGCCCTGGGCTGAGCTATGATATGATAAAGCCGAGAGATGGT 1080
Db 1021 GAAACACCTCTCAGGCTTAGCCCTGGGCTGAGCTATGATATGATAAAGCCGAGAGATGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTTGAATCTCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTTGAATCTCT 1140
Qy 1141 GCCTGCTTCATCACAGCTCCTCCTGTGAAAGAGAGTAGGTTTCTATGAATGGACT 1200
```

Db 1141 GCTGCTTCATACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGACATGTTTTGCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGACATGTTTTGCTTCAGCTAGAAAACACA 1260
QY 1261 ATGT 1264
Db 1261 ATGT 1264

RESULT 9

ADJ88278
ID ADJ88278 standard; DNA; 1284 BP.
XX
AC
XX
DT 06-MAY-2004 (first entry)
XX Human WT rGAL-8 R (galactosidase) DNA.
DE
XX Galactosidase; GAL; gene therapy; lysosomal storage disease;
XX Fabry's disease; Gaucher's disease; human; gene; ds.
KW
XX
OS Homo sapiens.

XX Key Location/Qualifiers
FH 1. .1284
FT CDS /*tag= a
FT /product= "Human WT rGAL-8R protein"
XX

US2004016021-A1.

22-JAN-2004.

23-JUN-2003; 2003US-00602219.

26-FEB-1988; 88US-00160766.

26-FEB-1988; 88US-00160771.

15-JUL-1988; 88US-00219279.

17-FEB-1989; 89US-00310881.

05-MAY-1989; 89US-00347637.

08-JUN-1989; 89US-00363138.

22-OCT-1990; 90US-00600244.

16-JAN-1991; 91US-00641617.

26-JUL-1991; 91US-00737899.

01-AUG-1991; 91US-00739143.

31-JUL-1992; 92US-00923692.

30-DEC-1992; 92US-00997733.

29-DEC-1993; 93US-00176414.

19-JAN-1994; 94US-00184237.

14-OCT-1994; 94US-00324003.

21-MAY-1999; 99US-00316572.

26-JUL-2000; 2000US-00626127.

13-NOV-2001; 2001US-00993059.

(TURP/) TURPEN T H.

(POGU/) POGUE G P.

(ERWI/) ERWIN R L.

(GRIL/) GRILL L K.

Turpen TH, Pogue GP, Erwin RL, Grill LK;

WPI; 2004-108227/11.

P-PSDB; ADJ88279.

XX New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
XX Claim 1; SEQ ID NO 13; 71pp; English.

XX The invention relates to nucleotide encoding galactosidase (GAL). The

CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
CC DNA.

XX
SQ Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;

Query Match 99.8%; Score 1264; DB 12; Length 1284;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCTCGCGCTTGCCTTCGCTTCCTGGCC 60

Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCTCGCGCTTGCCTTCGCTTCCTGGCC 60

QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCGCT 120

Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCGCT 120

QY 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180

Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180

QY 181 GATTCTCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

Db 181 GATTCTCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 241 TGGAGGATGAGGTTATGAGTACCTTCGATTTGATGACTGTGGATGGCTCCCAAGA 300

Db 241 TGGAGGATGAGGTTATGAGTACCTTCGATTTGATGACTGTGGATGGCTCCCAAGA 300

QY 301 GATTTCAGAGGAGGAGCTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

Db 301 GATTTCAGAGGAGGAGCTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

QY 361 GCTAATTATGTTTCACAGCAAAAGGAGCTGAAGTAGGATTTATGACAGATGTGGAATAAA 420

Db 361 GCTAATTATGTTTCACAGCAAAAGGAGCTGAAGTAGGATTTATGACAGATGTGGAATAAA 420

QY 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATACCTAGCATTTGATGATGCCAGACCTTTGCT 480

Db 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATACCTAGCATTTGATGATGCCAGACCTTTGCT 480

QY 481 GACTGGGAGTAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

Db 481 GACTGGGAGTAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

QY 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGGATGGCAGAGCATTTGTGTAC 600

Db 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGGATGGCAGAGCATTTGTGTAC 600

QY 601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAGAAAGCCCAATTTATACAGAAATCCGA 660

Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAGAAAGCCCAATTTATACAGAAATCCGA 660

QY 661 CAGTACTGCAATCAGTGGGCAATTTTGTGATTTGATGATTTCTCGGAAAGTATAAAG 720

Db 661 CAGTACTGCAATCAGTGGGCAATTTTGTGATTTGATGATTTCTCGGAAAGTATAAAG 720

QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGTCGACCAAGG 780

Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGTCGACCAAGG 780

QY 781 GGTGGAATGACCCAGATATGTTAGTATGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840

Db 781 GGTGGAATGACCCAGATATGTTAGTATGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840

QY 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTTCTAATGACCTC 900

Db 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTTCTAATGACCTC 900

QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGATAAGACGTAATGTCATCAAT 960

Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGATAAGACGTAATGTCATCAAT 960

Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCGCATCAAT 960
 QY 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACTTTGAAGTGG 1020
 Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACTTTGAAGTGG 1020
 QY 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGAGCTATGATAAACCGGAGGAGATTGGT 1080
 Db 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGAGCTATGATAAACCGGAGGAGATTGGT 1080
 QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTCGGTAAGAGGAGTGCCTGTAATCCT 1140
 Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTCGGTAAGAGGAGTGCCTGTAATCCT 1140
 QY 1141 GCGTGTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
 Db 1141 GCGTGTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
 QY 1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTAGAAAACACA 1260
 Db 1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTAGAAAACACA 1260
 QY 1261 ATGT 1264
 Db 1261 ATGT 1264

RESULT 10

ADM48682

ID ADM48682 standard; DNA; 1284 BP.

XX AC

ADM48682;

XX DT

03-JUN-2004 (first entry)

XX DE

Human wild type rGAL-8R DNA.

XX KW

Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;

KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;

KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;

XX KW gene; ds.

XX OS

Homo sapiens.

XX FH

Key Location/Qualifiers

FT CDS

1..1284

/*tag= a

/product= "GAL-8R protein"

XX FT

US2004023281-A1.

XX PN

XX XX

05-FEB-2004.

XX PF

23-JUN-2003; 2003US-00602220.

XX PR

26-FEB-1988; 88US-00160766.

XX PR

26-FEB-1988; 88US-00160771.

XX PR

17-FEB-1989; 89US-00310881.

XX PR

22-OCT-1990; 90US-00600244.

XX PR

31-JUL-1992; 92US-00923692.

XX PR

30-DEC-1992; 92US-00997733.

XX PR

29-DEC-1993; 93US-00176414.

XX PR

19-JAN-1994; 94US-00184237.

XX PR

14-OCT-1994; 94US-00324003.

XX PR

21-MAY-1999; 99US-00316572.

XX PR

26-JUL-2000; 2000US-00626127.

XX PR

13-NOV-2001; 2001US-00993059.

XX PA

(TURP/) TURPEN T H.

PA (KUMA/) KUMAGAI M H.

PA (FOGU/) FOGUE G P.

PA (ERWI/) ERWIN R L.

PA (GRIL/) GRILL L K.

XX

PI Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;

XX

DR WPI; 2004-142650/14.

XX

DR P-PSDB; ADM48683.

XX

New alpha-galactosidase polypeptides, useful in producing recombinant lysosomal enzymes for the treatment of lysosomal storage diseases, such as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs disease.

XX

PS Disclosure; SEQ ID NO 13; 72pp; English.

XX

The present invention relates to novel galactosidase (Gal) proteins such as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions of the present invention are useful for producing recombinant lysosomal enzymes for enzyme replacement therapy for treating human and animal lysosomal storage diseases such as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs disease. The present sequence is a human wild type rGAL-8R DNA. This sequence comprises a human rGAL-8 DNA and ER retention signal DNA. This sequence is used in the exemplification of the invention.

XX

SQ Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;

Query Match

Best Local Similarity 99.8%; Score 1264; DB 12; Length 1284;

Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 ATGAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60

Db

1 ATGAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60

QY

61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGAGCGCT 120

Db

61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGAGCGCT 120

QY

121 ACCATGGGCTGGCTGACCTGGAGGCGCTTCATGTGCAACCTTTCATGCCAGGAAGGCCA 180

Db

121 ACCATGGGCTGGCTGACCTGGAGGCGCTTCATGTGCAACCTTTCATGCCAGGAAGGCCA 180

QY

181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGC 240

Db

181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGC 240

QY

241 TGGAAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTGGATGCTCCCAAGA 300

Db

241 TGGAAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTGGATGCTCCCAAGA 300

QY

301 GATTCAGAGGCGAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

Db

301 GATTCAGAGGCGAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

QY

361 GCTAATATGTTTACAGCAAGGAGCTGAAGCTAGGATTTATGCGAGATGTTGGAATAAA 420

Db

361 GCTAATATGTTTACAGCAAGGAGCTGAAGCTAGGATTTATGCGAGATGTTGGAATAAA 420

QY

421 ACCTCGCGAGGCTTCCCTGGGAGTTTGGATAGTACTAGCATTGATGCCAGACCTTTGCT 480

Db

421 ACCTCGCGAGGCTTCCCTGGGAGTTTGGATAGTACTAGCATTGATGCCAGACCTTTGCT 480

QY

481 GACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTTACTGTGACAGATTTGGAAATTTG 540

Db

481 GACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTTACTGTGACAGATTTGGAAATTTG 540

QY

541 GCAGATGTTTATAAGCAGATGCTTGGCCCTGATAGACTGGCGAGAGCATTTGTGTAC 600

Db

541 GCAGATGTTTATAAGCAGATGCTTGGCCCTGATAGACTGGCGAGAGCATTTGTGTAC 600

QY

601 TCCTGTGAGTGGCCCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660

Db

601 TCCTGTGAGTGGCCCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA 660

Db 541 GCAGATGGTTAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAAGCAATTTGTATC 600
 Qy 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
 Db 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
 Qy 661 CAGTACTGCAATACACTGGCGAAATTTTGTGTCGACATTTGATGATTTCTTGGAAAGTATAAG 720
 Db 661 CAGTACTGCAATACACTGGCGAAATTTTGTGTCGACATTTGATGATTTCTTGGAAAGTATAAG 720
 Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTTCTTGGACAGG 780
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTTCTTGGACAGG 780
 Qy 781 GGTGGAATGACCCAGATATTTAGTATTTGGCACTTTTGGCCCTCAGCTGGAAATCAGCAA 840
 Db 781 GGTGGAATGACCCAGATATTTAGTATTTGGCACTTTTGGCCCTCAGCTGGAAATCAGCAA 840
 Qy 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTTATATCATGCTTAATGACCTC 900
 Db 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTTATATCATGCTTAATGACCTC 900
 Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAATTTGCCATCAAT 960
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAATTTGCCATCAAT 960
 Qy 961 CAGGACCCCTTGGCAAGCAAGGTTACACAGCTTTAGACAGGAGACAACTTTGAAGTGTG 1020
 Db 961 CAGGACCCCTTGGCAAGCAAGGTTACACAGCTTTAGACAGGAGACAACTTTGAAGTGTG 1020
 Qy 1021 GAACGACCTCTCTCAGCTTAGCTGGCTGTAGCTATGATATAAACCAGGAGATTTGT 1080
 Db 1021 GAACGACCTCTCTCAGCTTAGCTGGCTGTAGCTATGATATAAACCAGGAGATTTGT 1080
 Qy 1081 GGACCTCGCTTATACCATCCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCT 1140
 Db 1081 GGACCTCGCTTATACCATCCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCT 1140
 Qy 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTTTCTATGAATGGACT 1200
 Db 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTTTCTATGAATGGACT 1200
 Qy 1201 TCAAGTTAAGAGTACATAAATCCACAGGACCTGTTTTCTTCAGCTAGAGAAACACA 1260
 Db 1201 TCAAGTTAAGAGTACATAAATCCACAGGACCTGTTTTCTTCAGCTAGAGAAACACA 1260
 Qy 1261 ATGT 1264
 Db 1261 ATGT 1264

RESULT 12

AEA27446
 ID AEA27446 standard; DNA; 1284 BP.

XX AEA27446;

AC AEA27446;

XX 11-AUG-2005 (first entry)

DE Human alpha-galactosidase DNA, rGAL-8R, SEQ ID NO: 13.

XX Gauchers disease; metabolic; neurological disease; niemann pick disease;
 KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
 KW antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
 KW Gene; db.

OS Homo sapiens.

XX Key Location/Qualifiers

FT 1..1284

FT /tag= a

FT /product= "Human alpha-galactosidase protein"

XX PN US2005125859-A1.
 XX PD 09-JUN-2005.
 XX PF 08-NOV-2004; 2004US-00984389.
 XX PR 26-JUL-2000; 2000US-00626127.
 XX PR 13-NOV-2001; 2001US-0093059.
 XX PR 20-MAR-2002; 2002US-00103327.
 XX (LARG-) LARGE SCALE BIOLOGY CORP.
 XX PA Garger SJ, Turpen TH, Kumagai MH;
 XX PI P-PSDB; AEA27447.
 XX DR New isolated polypeptides useful for producing lysosomal enzymes in
 XX PT plants to be utilized in enzyme replacement therapy or for the
 XX PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
 XX PT Gaucher's disease.
 XX PS Disclosure; SEQ ID NO 13; 88pp; English.
 XX CC The present invention relates to the production of human and animal
 XX CC lysosomal enzymes in plants by a transient plant expression system. The
 XX CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
 XX CC galactosidase (Gal) enzymes having a post-translational modification
 XX CC provided by the plant expression system. The invention is useful in
 XX CC enzyme replacement therapy for treating lysosomal storage diseases such
 XX CC as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
 XX CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
 XX CC also useful in researches for developing new approaches to medical
 XX CC treatment of lysosomal storage diseases and in industrial processes
 XX CC involving enzymatic substrate hydrolysis. The present sequence is the
 XX CC human alpha-galactosidase DNA.
 XX SQ Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;

Query Match 99.8%; Score 1264; DB 14; Length 1284;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ATGAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCGCTTCTTGGCC 60
 Db 1 ATGAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCGCTTCTTGGCC 60
 Qy 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
 Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
 Qy 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCACTTGGCTGAGGAGGAGCA 180
 Db 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCACTTGGCTGAGGAGGAGCA 180
 Qy 181 GATTCTTCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
 Db 181 GATTCTTCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
 Qy 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300
 Db 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300
 Qy 301 GATTGAGGAGGAGACTTTCAGGACAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGTA 360
 Db 301 GATTGAGGAGGAGACTTTCAGGACAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGTA 360
 Qy 361 GCTAATTATGTTTACAGCAAGGAGCTGAAGCTAGGAGTTTATGCAAGATGTTGGAATAAA 420
 Db 361 GCTAATTATGTTTACAGCAAGGAGCTGAAGCTAGGAGTTTATGCAAGATGTTGGAATAAA 420
 Qy 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATATACGACATTTGATGCCAGACCTTTGCT 480

Db 361 GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCGATGTTGGAAATAAA 420
Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATATACGACATTTGATGCGCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATATACGACATTTGATGCGCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTCTCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAAATTTG 540
Db 481 GACTGGGAGTAGATCTCTCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAAATTTG 540
Qy 541 GCAGATGTTTAAAGCACAATGTCCTTGTGCGCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Db 541 GCAGATGTTTAAAGCACAATGTCCTTGTGCGCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCCCTTTTATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCCCTTTTATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGCGGAAATTTTGTGCTGACATTTGATGATTTCTGGAAGATATAAG 720
Db 661 CAGTACTGCAATCACTGCGGAAATTTTGTGCTGACATTTGATGATTTCTGGAAGATATAAG 720
Qy 721 AGTATCTTGGACTGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db 721 AGTATCTTGGACTGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACAGGG 780
Qy 781 GGTGGAATGACCCAGATATGTTAGTGATTTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTGATTTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGTAATGGCCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGTAATGGCCATCAAT 960
Qy 961 CAGACCCCTTGGGCAAGCAAGGTTACCACTTACAGCTTACAGAGGAGCAACTTTGNAAGTGG 1020
Db 961 CAGACCCCTTGGGCAAGCAAGGTTACCACTTACAGCTTACAGAGGAGCAACTTTGNAAGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGAGGCTGTAGCTATGATAAACCAGCAGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGAGGCTGTAGCTATGATAAACCAGCAGAGATTGGT 1080
Qy 1081 GGACCTGCTCTTATACCATGCGAGTCTCTTCCCTGAGGTAAGGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTGCTCTTATACCATGCGAGTCTCTTCCCTGAGGTAAGGAGTGGCTGTAATCCT 1140
Qy 1141 GCCTGCTTATCACACAGCTCTCTCCTGTGAAAAGGAGTAGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTATCACACAGCTCTCTCCTGTGAAAAGGAGTAGGCTTCTATGAATGGACT 1200
Qy 1201 TCAAGTTAAGAGTACATAAATCCCAAGGCACTGTTTGTCTTCAAGTACAGAAACACA 1260
Db 1201 TCAAGTTAAGAGTACATAAATCCCAAGGCACTGTTTGTCTTCAAGTACAGAAACACA 1260
Qy 1261 ATGTA 1265
Db 1261 ATGCA 1265

RESULT 14
ADD84746
ID ADD84746 standard; DNA; 1278 BP.
XX
AC ADD84746;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human alpha-galactosidase rGAL-4 DNA.
XX

KW Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..1278
FT /*tag= a
FT /product= "Human rGAL-4"
XX
PN US2003106095-A1.
XX
PD 05-JUN-2003.
XX
PF 20-MAR-2002; 2002US-00103327.
XX
PR 26-JUL-2000; 2000US-00626127.
XX
PR 13-NOV-2001; 2001US-00993059.
XX
XX (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2003-801257/75.
DR P-PSDB; ADD84747.
XX
XX
PT New polynucleotide for producing active recombinant human and animal
PT lysosomal enzymes in a plant expression system that can be used in enzyme
PT replacement therapy.
XX
XX Claim 1; SEQ ID NO 7; 77pp; English.
XX
XX The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for
CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents DNA encoding a human alpha-
CC galactosidase derivative polypeptide of the invention.
XX
SQ Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;

Query Match 99.8%; Score 1263.4; DB 10; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTGCTTCTGSCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTGCTTCTGSCC 60
Qy 61 CTGCTTCTGGGACATCCCTGGGCTAGACACTGGACAATGGATTGGCAAGGACGCT 120
Db 61 CTGCTTCTGGGACATCCCTGGGCTAGACACTGGACAATGGATTGGCAAGGACGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCACTGCAACCTTGAATGGCAAGAGCCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCACTGCAACCTTGAATGGCAAGAGCCA 180
Qy 181 GATTCTGTCATCAGTGAGAAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCTGTCATCAGTGAGAAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATTTGATGATGCTGGTGGCTTCCCAAGA 300
Db 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATTTGATGATGCTGGTGGCTTCCCAAGA 300
Qy 301 GATTCAAGGAGAGCTTTCAGGAGAGCCCTCAGGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAAGGAGAGCTTTCAGGAGAGCCCTCAGGAGAGCCCTCAGGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTATGTTTCACAGCAAGGAGCTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420

Db 361 GCTAATTATGTTACAGCAAGAGCTGAAGCTAGGATTTATGCGATGTTGGAATAAA 420
Qy 421 ACTCGGAGGCTTCCCTGGGAGTTTGGATAGTATGAGCATTTGATGCCAGACCTTTGCT 480
Db 421 ACTCGGAGGCTTCCCTGGGAGTTTGGATAGTATGAGCATTTGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGCTAAAATTTGATGCTTGTACTGTGACAGTTTGGAAAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGCTTGTACTGTGACAGTTTGGAAAATTG 540
Qy 541 CGAGATGTTTAAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTAC 600
Db 541 CGAGATGTTTAAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTAC 600
Qy 601 TCTGTAGTGGCTCTTTATATGATGGCCCTTTCAAAGCCCAATATACAGAAATCGA 660
Db 601 TCTGTAGTGGCTCTTTATATGATGGCCCTTTCAAAGCCCAATATATACAGAAATCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTTGTGATGATTTGCTGCAATTTGCTGGAATAAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGATGATTTGCTGCAATTTGCTGGAATAAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATTTGCTGGACAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATTTGCTGGACAGG 780
Qy 781 GGTGGATGACCCAGATATGTTAGTATGTCGCACTTTGGCCTCAGCTGGAAATCAGAA 840
Db 781 GGTGGATGACCCAGATATGTTAGTATGTCGCACTTTGGCCTCAGCTGGAAATCAGAA 840
Qy 841 GTAACCTCAGATGGCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGTAAATTCGCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGTAAATTCGCATCAAT 960
Qy 961 CAGGACCCCTGGGAGCAAGGTTACAGCTTAGACGGGAGACAACTTGAAGTGG 1020
Db 961 CAGGACCCCTGGGAGCAAGGTTACAGCTTAGACGGGAGACAACTTGAAGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGATGATATAACCGGAGGAGATTGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGATGATATAACCGGAGGAGATTGT 1080
Qy 1081 GGACCTCGCTTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAATCCT 1140
Qy 1141 GCTGCTTATCACAGCTCTCCCTGTGAAAGAGAGTGGCTTCTATGAATGGACT 1200
Db 1141 GCTGCTTATCACAGCTCTCCCTGTGAAAGAGAGTGGCTTCTATGAATGGACT 1200
Qy 1201 TCAAGTTAAGAGTACATAAATCCACAGGCTGTTTTCCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGTTAAGAGTACATAAATCCACAGGCTGTTTTCCTTCAGCTAGAAAACACA 1260
Qy 1261 ATGTA 1265
Db 1261 ATGCA 1265

RESULT 15
ID ADJ88272
XX ADJ88272 standard; DNA; 1278 BP.
XX
AC ADJ88272;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human rGAL- 4 (galactosidase) DNA.

XX Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease; human; gene; ds.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
CDS 1. 1278
FT /*tag= a
FT /product= "Human WT rGAL-4 protein"
FT /partial
FT /note= "No start codon"
XX
PN US2004016021-A1.
XX
PD 22-JAN-2004.
XX
PF 23-JUN-2003; 2003US-00602219.
XX
PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1993; 93US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX (TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
PI Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-108227/11.
DR P-PSDB; ADJ88273.
XX
PT New lysosomal enzymes, useful in treating human and animal lysosomal storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
PS Claim 1; SEQ ID NO 7; 71pp; English.
XX
CC The invention relates to nucleotide encoding galactosidase (GAL). The invention is useful in gene therapy. The polynucleotides and polypeptides are useful in treating human and animal lysosomal storage diseases, e.g. Fabry's disease and Gaucher's diseases. The present sequence is human GAL DNA.
XX
SQ Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;
Query Match 99.8%; Score 1263.4; DB 12; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTCGCGCTTGCCTTCCTCGGCC 60
Db 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTCGCGCTTGCCTTCCTCGGCC 60
Qy 61 CTCGTTTCTGGACATCCCTGGGCTAGACACTGGACATGGATTGGCAAGGACGCT 120
Db 61 CTCGTTTCTGGACATCCCTGGGCTAGACACTGGACATGGATTGGCAAGGACGCT 120

QY 121 ACCATGGGCTGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Db 121 ACCATGGGCTGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
QY 181 GATTCCTGCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
Db 181 GATTCCTGCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
QY 241 TGGAAAGGATGCAAGGTATGAGTACCTCTGCATTTGATGATGCTGTGGATGGCTCCCCCAAGA 300
Db 241 TGGAAAGGATGCAAGGTATGAGTACCTCTGCATTTGATGATGCTGTGGATGGCTCCCCCAAGA 300
QY 301 GATTCAGAAAGCAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAGAAAGCAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGGAATTTATGCAGATGTTGGAAATAAA 420
Db 361 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGGAATTTATGCAGATGTTGGAAATAAA 420
QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATACAGACATTTGATGCCAGACCTTTTGTCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATACAGACATTTGATGCCAGACCTTTTGTCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGATTTGGAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGATTTGGAAATTTG 540
QY 541 GCAGATGGTTAAGCAGATGCTTGGCCCTGAATAGACTGGCAGAGCAATTTGTGTAC 600
Db 541 GCAGATGGTTAAGCAGATGCTTGGCCCTGAATAGACTGGCAGAGCAATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATTTGATGATTCCTGGAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATTTGATGATTCCTGGAAAGTATAAG 720
QY 721 AGTATCTTGGACTGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGCACAGGG 780
Db 721 AGTATCTTGGACTGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGCACAGGG 780
QY 781 GGTTCGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Db 781 GGTTCGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTTATTCATGTCTAAAGCCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTTATTCATGTCTAAAGCCTC 900
QY 901 CGACATCATGCGCTCAAGCCAAAGCTCTCCTTCAGGATAGGAGTAATTCCTCAATCAAT 960
Db 901 CGACATCATGCGCTCAAGCCAAAGCTCTCCTTCAGGATAGGAGTAATTCCTCAATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATGCGAGTGTCTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
Db 1081 GGACCTCGCTCTTATACCATGCGAGTGTCTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCTCCTGTGAAAGAGAGTAGGGTCTTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCTCCTGTGAAAGAGAGTAGGGTCTTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAAGTCACATAAATCCCAAGGCACTGTTTTGCTTTCAGCTAGAAAAACACA 1260

Db 1201 TCAAGGTTAAGAAGTCACATAAATCCCAAGGCACTGTTTTGCTTTCAGCTAGAAAAACACA 1260
QY 1261 ATGTA 1265
Db 1261 ATGCA 1265

Search completed: December 26, 2005, 05:23:09
Job time : 775 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 04:17:59 ; Search time 4817 Seconds
(without alignments)
12296.543 Million cell updates/sec

Title: US-10-602-219-11
Perfect score: 1266
Sequence: 1 atgcagctgaggaaacccaga.....agctagaaacacaaatgtaa 1266

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_est3:*
- 4: gb_hic:*
- 5: gb_est4:*
- 6: gb_est5:*
- 7: gb_est6:*
- 8: gb_est7:*
- 9: gb_gss1:*
- 10: gb_gss2:*
- 11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1261.8	99.7	1277	4	CR605654	CR605654 full-leng
2	1258.4	99.4	1266	4	CR607242	CR607242 full-leng
3	1248	98.6	1253	4	CR617861	CR617861 full-leng
4	1169.8	92.4	1290	10	AY408540	AY408540 Homo sapi
5	1149.4	90.8	1290	10	AY408541	AY408541 Pan trogl
6	968.6	76.5	1086	1	AL554978	AL554978 AL554978
7	926.8	73.2	1014	5	BX354096	BX354096 BX354096
8	922.8	72.9	999	1	AL552630	AL552630 AL552630
9	915.8	72.3	1067	3	BM564282	BM564282 AGENCOURT
10	915	72.3	1026	1	AL577581	AL577581 AL577581
11	899.8	71.1	1376	4	AK054547	AK054547 Mus muscu
12	899.8	71.1	2962	4	AK040081	AK040081 Mus muscu
13	890	70.3	1133	7	CO645672	CO645672 ILLUMIGEN
14	883.2	69.8	922	7	CO645623	CO645623 ILLUMIGEN
15	880.8	69.6	1071	1	AL575861	AL575861 AL575861
16	874.6	69.1	1005	3	BQ062192	BQ062192 AGENCOURT
17	870	68.7	984	7	CO646251	CO646251 ILLUMIGEN
18	862.4	68.1	920	7	CO645464	CO645464 ILLUMIGEN
19	856.4	67.6	923	7	CT005156	CT005156 CT005156
20	828.6	65.5	1296	10	AY408542	AY408542 Mus muscu
21	814.2	64.3	976	5	BQ956043	BQ956043 AGENCOURT
22	807.8	63.8	937	5	BQ934640	BQ934640 AGENCOURT

23	797	63.0	922	5	BU540848	BU540848 AGENCOURT
24	794.8	62.8	898	5	BU191867	BU191867 AGENCOURT
25	783.8	61.9	888	5	BU154569	BU154569 AGENCOURT
26	782	61.8	852	6	CA454083	CA454083 AGENCOURT
27	768	60.7	801	5	BU596617	BU596617 AGENCOURT
28	765.4	60.5	927	2	BE622583	BE622583 601440703
29	765.2	60.4	813	5	EX344841	EX344841 BX344841
30	761.4	60.1	871	6	CA454143	CA454143 AGENCOURT
31	757.6	59.8	1025	3	BM450649	BM450649 AGENCOURT
32	752.2	59.4	958	6	CA487415	CA487415 AGENCOURT
33	747.6	59.1	849	6	CA487531	CA487531 AGENCOURT
34	747	59.0	796	8	DR156332	DR156332 HESC2 64
35	746.8	59.0	845	7	CR983967	CR983967 CR983967
36	742.6	58.7	908	2	BE379425	BE379425 601237275
37	742.6	58.7	782	2	BI224248	BI224248 602940538
38	742	58.6	871	3	BQ225444	BQ225444 AGENCOURT
39	741.4	58.6	1066	2	BG824323	BG824323 602727346
40	741.2	58.5	921	6	CA488935	CA488935 AGENCOURT
41	737.2	58.2	879	2	BG824387	BG824387 602728627
42	729	57.6	749	5	BX344842	BX344842 BX344842
43	727.6	57.5	857	5	BX374627	BX374627 BX374627
44	713.8	56.4	800	6	CB529027	CB529027 UI-H-FT2-
45	698.8	55.2	767	3	BI753664	BI753664 603023559

ALIGNMENTS

RESULT 1
CR605654
LOCUS full-length cDNA clone CS0DB008YB11 of Neuroblastoma Cot
DEFINITION 10-normalized of Homo sapiens (human).
ACCESSION CR605654.1 GI:50486461
VERSION HTC; CNSLT_CDNA.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1277)
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Paraday Avenue
REFERENCE 2 (bases 1 to 1277)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequençage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source Location/Qualifiers
1..1277
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DB008YB11"
/tissue_type="Neuroblastoma Cot 10-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 99.7%; Score 1261.8; DB 4; Length 1277;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1263; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCGCTTCTGGCC 60
Db |||||
QY 10 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCGCTTCTGGCC 69
Db |||||
QY 61 CTCGTTTCTGGGACATCCTCGGGCTAGAGCACTGGCAATGGAATGGCAAGGAGCGCT 120
Db |||||
QY 70 CTCGTTTCTGGGACATCCTCGGGCTAGAGCACTGGCAATGGAATGGCAAGGAGCGCT 129
Db |||||
QY 121 ACCATGGGTGGCTGATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCA 180
Db |||||
QY 130 ACCATGGGTGGCTGATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCA 189
Db |||||
QY 181 GATTCCTGATCAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db |||||
QY 190 GATTCCTGATCAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 249
Db |||||
QY 241 TGGAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCCAAGA 300
Db |||||
QY 250 TGGAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCCAAGA 309
Db |||||
QY 301 GATTCAGAGGAGACTTCAGGAGACCTCAGCGCTTTCCTCATGCGGATTCGCCAGCTA 360
Db |||||
QY 310 GATTCAGAGGAGACTTCAGGAGACCTCAGCGCTTTCCTCATGCGGATTCGCCAGCTA 369
Db |||||
QY 361 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGAGTTATGCAAGATGTTGGAATAAA 420
Db |||||
QY 370 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGAGTTATGCAAGATGTTGGAATAAA 429
Db |||||
QY 421 ACCTGGCAGGCTTCCTCGGGAGTTTGGATACTACGACATGATGCCAGACCTTTGCT 480
Db |||||
QY 430 ACCTGGCAGGCTTCCTCGGGAGTTTGGATACTACGACATGATGCCAGACCTTTGCT 489
Db |||||
QY 481 GACTGGGAGTAGACTGCTGAAAATTTGATGTTGTTACTGTGACAGTTTGGAAAATTG 540
Db |||||
QY 490 GACTGGGAGTAGACTGCTGAAAATTTGATGTTGTTACTGTGACAGTTTGGAAAATTG 549
Db |||||
QY 541 GCAGATGTTTAAAGCACATGCTCTGGCCCTGATAGGACTGGCAGAGAGCTTGTGTC 600
Db |||||
QY 550 GCAGATGTTTAAAGCACATGCTCTGGCCCTGATAGGACTGGCAGAGAGCTTGTGTC 609
Db |||||
QY 601 TCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db |||||
QY 610 TCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 669
Db |||||
QY 661 CAGTACTGCAATCAGCTGGGAAAATTTGCTGACATGATGATCTCTGGAAAAGATTAAG 720
Db |||||
QY 670 CAGTACTGCAATCAGCTGGGAAAATTTGCTGACATGATGATCTCTGGAAAAGATTAAG 729
Db |||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTCTGACACAGG 780
Db |||||
QY 730 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTCTGACACAGG 789
Db |||||
QY 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCTCAGCTGGAATCAGCAA 840
Db |||||
QY 790 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCTCAGCTGGAATCAGCAA 849
Db |||||
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATCATGTCTAATGACCTC 900
Db |||||
QY 850 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATCATGTCTAATGACCTC 909
Db |||||
QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGCTAATTTGCCATCAAT 960
Db |||||
QY 910 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGCTAATTTGCCATCAAT 969
Db |||||
QY 961 CAGGACCCCTTGGGCAAGCAAGGATACCACTTTAGACAGGAGCAACCTTTGAAGTGTGG 1020
Db |||||
QY 970 CAGGACCCCTTGGGCAAGCAAGGATACCACTTTAGACAGGAGCAACCTTTGAAGTGTGG 1029
Db |||||
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTTATGATAAACCCGAGGAGATTGT 1080
Db |||||
QY 1030 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTTATGATAAACCCGAGGAGATTGT 1089
Db |||||
QY 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140

Db 1090 GGACCTCGCTCTTATACCATCGCAGTTCCTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1149
QY 1141 GCCTGCTTATCACAACAGCTCCTCCTGTGAAAAGAAAGCTAGGTTCTTATGAATGGACT 1200
Db |||||
QY 1150 GCCTGCTTATCACAACAGCTCCTCCTGTGAAAAGAAAGCTAGGTTCTTATGAATGGACT 1209
Db |||||
QY 1201 TCAGGTTTAAAGTTCACATAAATCCACAGGCACTGTTTGCCTTCAGCTAGAAAACACA 1260
Db |||||
QY 1210 TCAGGTTTAAAGTTCACATAAATCCACAGGCACTGTTTGCCTTCAGCTAGAAAATACA 1269
Db |||||
QY 1261 ATGTA 1265
Db |||||
QY 1270 ATGCA 1274
Db |||||
RESULT 2
CR607242 1266 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0D1067YJ01 of Placenta Cot 25-normalized
DEFINITION of Homo sapiens (human).
ACCESSION CR607242
VERSION CR607242.1 GI:50488049
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1266)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
REFERENCE 2 (bases 1 to 1266)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
source
1..1266
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1067YJ01"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 99.4%; Score 1258.4; DB 4; Length 1266;
Best Local Similarity 99.9%; Pred.No. 0;
Matches 1259; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCGCTTCTGGCC 60
Db |||||
QY 7 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCGCTTCTGGCC 66
Db |||||
QY 61 CTCGTTTCTCGGGACATCCTCGGGCTAGAGCACTGGCAATGGAATGGCAAGGAGCGCT 120
Db |||||
QY 67 CTCGTTTCTCGGGACATCCTCGGGCTAGAGCACTGGCAATGGAATGGCAAGGAGCGCT 126
Db |||||
QY 121 ACCATGGGTGGCTGATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCA 180
Db |||||
QY 127 ACCATGGGTGGCTGATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCA 186
Db |||||

```
QY 181 GATTCTGTCATCAGTGGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Db 187 GATTCTGTCATCAGTGGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 246
QY 241 TCGAAGGATGCAAGGTTATGAGTACCTCTGCAATGATGATGCTGGTGGTCCCAAGA 300
Db 247 TCGAAGGATGCAAGGTTATGAGTACCTCTGCAATGATGATGCTGGTGGTCCCAAGA 306
QY 301 GATTGAGAGGAGAGCTTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 307 GATTGAGAGGAGAGCTTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 366
QY 361 GCTAATATATGTTACAGCAAGAGAGCTGAAGCTAGGATTTATGAGAGATGTTGGAATATAA 420
Db 367 GCTAATATATGTTACAGCAAGAGAGCTGAAGCTAGGATTTATGAGAGATGTTGGAATATAA 426
QY 421 ACCTGGCAGGCTCCCTGGGAGTTTGGATACCTAGCATTGATGCGCCAGACCTTTCCT 480
Db 427 ACCTGGCAGGCTCCCTGGGAGTTTGGATACCTAGCATTGATGCGCCAGACCTTTCCT 486
QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGATTTGGAATTTG 540
Db 487 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGATTTGGAATTTG 546
QY 541 GCAGATGTTTAAAGCAGATGCTTGGCCCTGAATGAGACTGGCAGAGCAATTTGTATC 600
Db 547 GCAGATGTTTAAAGCAGATGCTTGGCCCTGAATGAGACTGGCAGAGCAATTTGTATC 606
QY 601 TCTGTGAGTGCCCTCTTATATGTTGGCCCTTCAAAGCCCAATTTACAGAAATCCGA 660
Db 607 TCTGTGAGTGCCCTCTTATATGTTGGCCCTTCAAAGCCCAATTTACAGAAATCCGA 666
QY 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATTTGATGATTCCTGGAAAGATATAAG 720
Db 667 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATTTGATGATTCCTGGAAAGATATAAG 726
QY 721 AGTATCTTGGATGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACCAAGG 780
Db 727 AGTATCTTGGATGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACCAAGG 786
QY 781 GATTGGATGACCCAGATATGTTAGTATGTTGCAATTTGCTGCAATTTGATGATTCCTGGAAAGATATAAG 840
Db 787 GATTGGATGACCCAGATATGTTAGTATGTTGCAATTTGCTGCAATTTGATGATTCCTGGAAAGATATAAG 846
QY 841 GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
Db 847 GTAACCTCAGATGCGCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTTAATGACCTC 906
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAAT 960
Db 907 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAAT 966
QY 961 CAGGACCCCTTGGGAGCAAGGTTACAGCTTTAGACAGGAGAGACAACTTTGAAGTGTGG 1020
Db 967 CAGGACCCCTTGGGAGCAAGGTTACAGCTTTAGACAGGAGAGACAACTTTGAAGTGTGG 1026
QY 1021 GAACGACCTCTCTCAGGCTTACGCTGGGCTGATGATTAACCCGCGAGAGATTTGTT 1080
Db 1027 GAACGACCTCTCTCAGGCTTACGCTGGGCTGATGATTAACCCGCGAGAGATTTGTT 1086
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTCGGTTAAAGGAGTGGCTGTAACTCT 1140
Db 1087 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTCGGTTAAAGGAGTGGCTGTAACTCT 1146
QY 1141 GCCTGTTTATCAGCAGCTCTCTCTGTTGAAAGAGAGCTAGGTTTCTATGAATGGACT 1200
Db 1147 GCCTGTTTATCAGCAGCTCTCTCTGTTGAAAGAGAGCTAGGTTTCTATGAATGGACT 1206
QY 1201 TCAAGGTTAAGAGTACATTAATCCCAAGGACCTGTTTGTCTTCAGCTAGAAACACA 1260
Db 1207 TCAAGGTTAAGAGTACATTAATCCCAAGGACCTGTTTGTCTTCAGCTAGAAATACA 1266
```

```
RESULT 3
CR617861
LOCUS
DEFINITION
full-length cDNA clone CS0DK009YB12 of HeLa cells Cot 25-normalized
of Homo sapiens (human).
ACCESSION
CR617861
VERSION
CR617861.1 GI:50498668
KEYWORDS
HTC; CNSLT cDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1. (bases 1 to 1253)
AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
REMARK
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
Genoscope.
2. (bases 1 to 1253)
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
source
1. 1253
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK009YB12"
/tissue type="HeLa cells Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 98.6%; Score 1248; DB 4; Length 1253;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTCGCTTCTGGCC 60
Db 6 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTCGCTTCTGGCC 65
QY 61 CTGCTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120
Db 66 CTGCTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 125
QY 121 ACCATGGCTGGCTGCACTGGGAGGCTTTCATGTGCAACCTTTGACTGCCAGGAAGAGCCA 180
Db 126 ACCATGGCTGGCTGCACTGGGAGGCTTTCATGTGCAACCTTTGACTGCCAGGAAGAGCCA 185
QY 181 GATTCCTGCATCAGTGAGAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 186 GATTCCTGCATCAGTGAGAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 245
QY 241 TGGAGGATGAGAGGTTATGATGATCTCTGCAATTTGATGATCTGTTGATGGCTCCCAAGA 300
Db 246 TGGAGGATGAGAGGTTATGATGATCTCTGCAATTTGATGATCTGTTGATGGCTCCCAAGA 305
QY 301 GATTGAGAGGAGAGCTTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 306 GATTGAGAGGAGAGCTTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 365
QY 361 GCTAATATGTTACAGCAAGAGAGCTGAAGCTAGGATTTATGAGATGTTGGAATATAA 420
Db 366 GCTAATATGTTACAGCAAGAGAGCTGAAGCTAGGATTTATGAGATGTTGGAATATAA 425
```


QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTGGAAGATATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTGGAAGATATAAG 720
QY 721 AGTATCTGGAGTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTTGCTGGACAGGG 780
Db 721 AGTATCTGGAGTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTTGCTGGACAGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTTGCACTTTGGCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTATGTTGCACTTTGGCTCAGCTGGAATCAGCAA 840
QY 841 GTAACATGAGTGGCCCTCTGGGCTATCATGCTGCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACATGAGTGGCCCTCTGGGCTATCATGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 901 CGACATCATGCGCTCAAGCCAAAGCTCTCTTCAAGATAGGAGTAAATGCGCATCAAT 960
Db 901 CGACATCATGCGCTCAAGCCAAAGCTCTCTTCAAGATAGGAGTAAATGCGCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTACAGAGGAGAGAACTTTGAACTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTACAGAGGAGAGAACTTTGAACTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAAACCGGACAGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAAACCGGACAGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACATGCACTGCTTCCCTGGGTAAAGAGTGGCCCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACATGCACTGCTTCCCTGGGTAAAGAGTGGCCCTGTAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAGAGCTAGGCTTATGAAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAGAGCTAGGCTTATGAAATGGACT 1200
QY 1201 TCAAGTTAAGAGTACATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAGAAACACA 1260
Db 1201 TCAAGTTAAGAGTACATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAGAAATACA 1260
QY 1261 ATGTA 1265
Db 1261 ATGCA 1265

RESULT 5

AY408541
LOCUS
DEFINITION
Pan troglodytes GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION
AY408541
VERSION
AY408541.1
KEYWORDS
GI:39764512
SOURCE
Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Pan.
REFERENCE
1 (bases 1 to 1290)
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
JOURNAL
PUBMED
14671302
REFERENCE
2 (bases 1 to 1290)
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.

TITLE
JOURNAL
COMMENT
FEATURES
source
gene
ORIGIN
Query Match
Best Local Similarity
Matches 1153; Conservative
QY 1
Db 1
QY 61
Db 61
QY 121
Db 121
QY 181
Db 181
QY 241
Db 241
QY 301
Db 301
QY 361
Db 361
QY 421
Db 421
QY 481
Db 481
QY 541
Db 541
QY 601
Db 601
QY 661
Db 661
QY 721
Db 721
QY 781

Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
Location/Qualifiers
1..1290
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1290
/gene="GLA"
/locus_tag="HCM3258"

90.8%; Score 1149.4; DB 10; Length 1290;
91.1%; Pred. No. 0;
0; Mismatches 112; Indels 0; Gaps 0;

1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTGCTTCTGCGC 60
1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTGCTTCTGCGC 60
61 CTGCTTCTGGGACATCCCTTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120
61 CTGCTTCTGGGACATCCCTTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120
121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCACTGTGCAACCTTGTGCTGGCAAGGACGCA 180
121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCACTGTGCAACCTTGTGCTGGCAAGGACGCA 180
181 GATTCTCTGCATCAGTGAAGACTCTTCACTGAGATGGCAGAGCTCATGGTCTCNGAAGGC 240
181 GATTCTCTGCATCAGTGAAGACTCTTCACTGAGATGGCAGAGCTCATGGTCTCNGAAGGC 240
241 TGGAAAGATGAGGTTATGAGTACCTCTGCAATTGATGACTGTGGATGGCTCCCAAGA 300
241 TGGAAAGATGAGGTTATGAGTACCTCTGCAATTGATGACTGTGGATGGCTCCCAAGA 300
301 GATTGAGAGGACAGACTTTCAGGACAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
301 GATTGAGAGGACAGACTTTCAGGACAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
361 GCTAATTATGTTTACAGCAAGAGACTGAAGCTAGGATTTATGCAAGATGTTGGAAATAA 420
361 NNN 420
421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACATGACATGATGATGATGATGATGATGAT 480
421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACATGACATGATGATGATGATGATGATGAT 480
481 GACTGGGAGTAGATCTGCTAAATTTTCTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
481 GACTGGGAGTAGATCTGCTAAATTTTCTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
541 GCAGATGTTTATAAGACACATGCTCTTGGCCCTGAAATAGGACTGGCAGAGCAATGTTGTAC 600
541 GCAGATGTTTATAAGACACATGCTCTTGGCCCTGAAATAGGACTGGCAGAGCAATGTTGTAC 600
601 TCCTGTGAGTGGCTCTTTTATATGTTGGCCCTTCAAAGGCCAATATATACAGAAATCCGA 660
601 NNN 660
661 CAGTACTGCAATCACTGGCGAAATTTTCTGCTGACATGATGATGATGATGATGATGATGATG 720
661 CAGTACTGCAATCACTGGCGAAATTTTCTGCTGACATGATGATGATGATGATGATGATGATG 720
721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGTTGTTGTTGTTGTTGTT 780
721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGTTGTTGTTGTTGTTGTT 780
781 GGTGGAATGACCCAGATATGTTAGTATGTTGGCACTTGGCTTGGCTTGGCTTGGCTTGGCTT 840

Db 781 GGTGGAATGACCCAGATATGTTAGTATGTCGAACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTCAGATGCGCTCTGGGCTATCATGCGTCTCCCTTTATTCATGCTTAATCAGCTC 900
Db 841 GTAACCTCAGATGCGCTCTGGGCTATCATGCGTCTCCCTTTATTCATGCTTAATCAGCTC 900
Qy 901 GCACACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAGGATAAGGAGTAAATTCGCCATCAAT 960
Db 901 GCACACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAGGATAAGGAGTAAATTCGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGATACAGCTTACAGCTTACAGAGGAGACAACTTTGAAGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGATACAGCTTACAGCTTACAGAGGAGACAACTTTGAAGTGG 1020
Qy 1021 GAAACACCTCTCTCAGGCTTACGCTGGGCTGAGCTATGATTAACCGGACGAGATTGGT 1080
Db 1021 GAAACACCTCTCTCAGGCTTACGCTGGGCTGAGCTATGATTAACCGGACGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAACTCT 1140
Db 1081 GGACCTCGCTCTTATACCATCCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAACTCT 1140
Qy 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGGAAAGCTAGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGGAAAGCTAGGTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCACATTAATCCACAGGCACTGTTTCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCACATTAATCCACAGGCACTGTTTCTTCAGCTAGAAAATACA 1260
Qy 1261 ATGTA 1265
Db 1261 ATGCA 1265

RESULT 6
AL554978
LOCUS AL554978 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
DEFINITION AL554978 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
ACCESSION AL554978
VERSION AL554978.3 GI:45859723
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1086)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:31276787.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo (dt) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?e=CS0DK009BC06QP1&c=10506.r.
1. .1086
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK009YE12"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"

FEATURES
source

/clone.lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dt)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 76.5%; Score 968.6; DB 1; Length 1086;
Best Local Similarity 98.0%; Pred. No. 8e-279;
Matches 1012; Conservative 9; Mismatches 8; Indels 4; Gaps 4;

Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGTTTCGCTTCTTGCC 60
Db 6 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGTTTCGCTTCTTGCC 64
Qy 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGATGGCAAGAGCGCT 120
Db 65 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGATGGCAAGAGCGCT 124
Qy 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCGCAGGAAGGCA 180
Db 125 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCGCAGGAAGGCA 184
Qy 181 GATTCTGTCATCAGTGGAAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 185 GATTCTGTCATCAGTGGAAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 244
Qy 241 TGGGAAGATCAGAGTTATGATGACTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db 245 TGGGAAGATCAGAGTTATGATGACTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 304
Qy 301 GATTTCAGGAAGCAGACTTTCAGGAGACCTTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 305 GATTTCAGGAAGCAGACTTTCAGGAGACCTTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 364
Qy 361 GCTAATTTATGTTTCAGCAAAAGGACTGAAGCTAGGGAATTTATGCAAGATGTTGGAAATAAA 420
Db 365 GCTAATTTATGTTTCAGCAAAAGGACTGAAGCTAGGGAATTTATGCAAGATGTTGGAAATAAA 424
Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACATGACGATGATGCCAGACCTTTGCT 480
Db 425 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACATGACGATGATGCCAGACCTTTGCT 484
Qy 481 GACTGGGAGTAGATCTCTAAATTTTCATGCTGTTTACTGTGACAGTTTGGAAATTTG 540
Db 485 GACTGGGAGTAGATCTCTAAATTTTCATGCTGTTTACTGTGACAGTTTGGAAATTTG 544
Qy 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGTAATAGGACTGGCAGAGCAATTTGTAC 600
Db 545 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGTAATAGGACTGGCAGAGCAATTTGTAC 604
Qy 601 TCCTGTGAGTGGCTCTTTTATATGTCGCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 605 TCCTGTGAGTGGCTCTTTTATATGTCGCTTTCAAAAGCCCAATTATACAGAAATCCGA 664
Qy 661 CAGTACTGCAATCACTGCGGAAATTTTCTGACATTTGATGATTCCTGAAAGATATAAG 720
Db 665 CAGTACTGCAATCACTGCGGAAATTTTCTGACATTTGATGATTCCTGAAAGATATAAG 724
Qy 721 AGTATCTTTGGACTGGACATCTTTTAAACAGGAGAGAATTTGATGTTGCTGGACAGGG 780
Db 725 AGTATCTTTGGACTGGACATCTTTTAAACAGGAGAGAATTTGATGTTGCTGGACAGGG 784
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAAATCAGCAA 840
Db 785 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAAATCAGCAA 844
Qy 841 GTAACCTCAGATGCGCCTCTGGGCTA-TCATGGCTGCTCTTTTATTCATGCTCTAAAGACT 899
Db 845 GTAACCTCAGATGCGCCTCTGGGCTATTTCATGGCTGCTCTTTTATTCATGCTCTAAAGACT 904
Qy 900 CCGACACATCAGCCTTCAAGCCAAAGCTCTCTTCAGGATAGGACGCTAAATGTCATCA 959
|||||

Db 905 CCGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATATAGGAGCTWATTGGCCATCA 964

Qy 960 TCAGGACCCCTTTGGGCAACAAAGGTTACAGCTTATAGACAGGAGACAACTTTTGAAGTGTG 1019

Db 965 TCAGGA-CCCTTKGGCAACAAAGGTTA-CAGCTTAGAAGAGGCGRGACAWTTTGMAGTGTG 1022

Qy 1020 GGAACGACCTCTC 1032

Db 1023 GGAMGACTCTCTC 1035

RESULT 7
BX354096
LOCUS
DEFINITION BX354096 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DC014YH04 5-PRIME, mRNA sequence.
ACCESSION BX354096
VERSION BX354096.2 GI:46550107
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1014)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNAL On May 5, 2003 this sequence version replaced gi:30379845.
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, Cp 5706 - 91057 EVRY cedex - FRANCE
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?S=CS0DC014DD02QP1&c=10506.r.

FEATURES
Location/Qualifiers
1..1014
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC014YH04"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 73.2%; Score 926.8; DB 5; Length 1014;
Best Local Similarity 98.4%; Pred. No. 2.8e-266;
Matches 958; Conservative 8; Mismatches 5; Indels 3; Gaps 3;
Qy 11 GGAACCCAGAACTACATCTGGGCTGCGCGCTTCGCTTCCTGGCCCTCGTTTCT 70
Db 29 GGAACCCAGAACTACATCTGGGCTGCGCGCTTGC-STTCGCTTCCTGGCCCTCGTTTCT 87
Qy 71 GGGACATCCCTGGGCTAGACACTGGACAATGGATTGGCAAGGACGCTACCATGGCT 130
Db 88 GGGACATCCCTGGGCTAGACACTGGACAATGGATTGGCAAGGACGCTACCATGGCT 147
Qy 131 GGTGCTACCTGGGAGCGCTTCATGTGCAACTCTGACCTGCCAGGAAGCCAGATCTCTGCA 190
Db 148 GGTGCTACCTGGGAGCGCTTCATGTGCAACTCTGACCTGCCAGGAAGCCAGATCTCTGCA 207
Qy 191 TCAGTGAGAGCTCTTTCATGAGATGGCAGAGCTCATGCTTCAGAGGCTGGAAGGATG 250

Db 208 TCAGTGAGAGCTCTTTCATGAGATGGGAGAGCTCATGGTCTCAGAGGCTGGAAGGATG 267

Qy 251 CAGGTTATGAGTACCTCTCGCATTTGATGACTGTGTGATGGCTCCCAAGAGATTCAGAAG 310

Db 268 CAGGTTATGAGTACCTCTCGCATTTGATGACTGTGTGATGGCTCCCAAGAGATTCAGAAG 327

Qy 311 GCAGACTTCAGGAGACCCCTCAGCGCTTCTCATGGGATTCGCCAGCTAGCTAATTATG 370

Db 328 GCAGACTTCAGGAGACCCCTCAGCGCTTCTCATGGGATTCGCCAGCTAGCTAATTATG 387

Qy 371 TTCACAGCAAAAGGACTGAAGCTAGGATTTATGACAGATGTTGGAAATAAAACCTCGCAG 430

Db 388 TTCACAGCAAAAGGACTGAAGCTAGGATTTATGACAGATGTTGGAAATAAAACCTCGCAG 447

Qy 431 GCTTCCCTGGGAGTTTGGATACGACATTCAGTATGATGCCAGACCTTTGCTGACTGGGAG 490

Db 448 GCTTCCCTGGGAGTTTGGATACGACATTCAGTATGATGCCAGACCTTTGCTGACTGGGAG 507

Qy 491 TAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTGGCAGATGGT- 549

Db 508 TAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTGGCAGATGGT- 567

Qy 550 TATAGCACATCTCTCGCCCTGAATAGGACTGGCAGAGCATTTGTACTCTCTGTGAG 609

Db 568 TATAAGCACATCTCTCGCCCTGAATAGGACTGGCAGAGCATTTGTACTCTCTGTGAG 627

Qy 610 TGGCCTCTTTATATGTCGCCCTTTCAAAGGCCCAATTATACAGAAATCCGACAGTACTGC 669

Db 628 TGGCCTCTTTATATGTCGCCCTTTCAAAGGCCCAATTATACAGAAATCCGACAGTACTGC 687

Qy 670 AATCACTGGGAAATTTTCTGACATTCATGATGATTCCTGGAAGATTAAGAGATTCCTTG 729

Db 688 AATCACTGGGAAATTTTCTGACATTCATGATTCCTGGAAGATTAAGAGATTCCTTG 747

Qy 730 GACTGGACATCTTTTAAACAGGAGAGATTTGTGATGTTGCTGGACCGAGGGGTTGGAAT 789

Db 748 GACTGGACATCTTTTAAACAGGAGAGATTTGTGATGTTGCTGGACCGAGGGGTTGGAAT 807

Qy 790 GACCCAGATATGTTAGTATGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAAGTAACTCAG 849

Db 808 GACCCAGATATGTTAGTATGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAAGTAACTCAG 867

Qy 850 ATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTTAATGACCTCCGACACATC 909

Db 868 ATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTTAATGACCTCCGACACATC 927

Qy 910 AGCCCTCAAGCCCAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAATCAGGACCCC 969

Db 928 AGCCCTCAAGCCCAAGCTCTCTTCAGGATAGGACGTAATTTGCCA-CAAGCAGGACCCC 986

Qy 970 TTGGCAAGCAAGG 983

Db 987 TGGCMAGCAAGGK 1000

RESULT 8
AL552630
LOCUS AL552630 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0DI067YJ01 5-PRIME, mRNA sequence.
ACCESSION AL552630
VERSION AL552630.3 GI:45857411
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 999)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNAL On Feb 15, 2001 this sequence version replaced gi:31274445.

Contact: Genoscope
 Genoscope - Centre National de Sequencage
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 10506.r

For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?s=CS0DI067CE01QP1&c=10506.r.

FEATURES

Source

Location/Qualifiers
 1..999
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DI067YJ01"
 /issue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 72.9%; Score 922.8; DB 1; Length 999;
 Best Local Similarity 99.3%; Pred. No. 4.4e-265;
 Matches 976; Conservative 3; Mismatches 0; Indels 5; Gaps 5;
 QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGTTCCTCTCGGCC 60
 DB 7 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGTTCCTCTCGGCC 65
 QY 61 CTCGTTTCTCTGGGACATCCTCTGGGCTAGAGCACTGGACAATGATGTGCAAGGACGCT 120
 DB 66 CTCGTTTCTCTGGGACATCCTCTGGGCTAGAGCACTGGACAATGATGTGCAAGGACGCT 125
 QY 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTGTGACCTGCCAGGAGGCA 180
 DB 126 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTGTGACCTGCCAGGAGGCA 185
 QY 181 GATTCTCGATCAGTCAGGAAGCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240
 DB 186 GATTCTCGATCAGTCAGGAAGCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGC 245
 QY 241 TGAAGGATGCAAGTTATGATGACCTCTCATTTGATGATGATGATGATGATGATGATGATG 300
 DB 246 TGAAGGATGCAAGTTATGATGACCTCTCATTTGATGATGATGATGATGATGATGATGATG 305
 QY 301 GATTGAGGAGGAGCTTCAGGAGACCTCAGCGCTTCCCTCATGGGATTCGCCAGCTA 360
 DB 306 GATTGAGGAGGAGCTTCAGGAGACCTCAGCGCTTCCCTCATGGGATTCGCCAGCTA 365
 QY 361 GCTAATTATGTTTCAAGCAAGGAGCTGAAGCTAGGAGTTATGACATGTTGGAATAAA 420
 DB 366 GCTAATTATGTTTCAAGCAAGGAGCTGAAGCTAGGAGTTATGACATGTTGGAATAAA 425
 QY 421 ACTGCGGAGGCTTCCCTGGAGTTTGGATGATGATGATGATGATGATGATGATGATGATG 480
 DB 426 ACTGCGGAGGCTTCCCTGGAGTTTGGATGATGATGATGATGATGATGATGATGATGATG 485
 QY 481 GACTGGGAGTATGATCTGCTAAATTTGATGTTGTTTCTGTCGACATGTTGGAAATTTG 540
 DB 486 GACTGGGAGTATGATCTGCTAAATTTGATGTTGTTTCTGTCGACATGTTGGAAATTTG 545
 QY 541 GCAGATGTTTATAGACATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
 DB 546 GCAGATGTTTATAGACATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 605
 QY 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
 DB 606 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 665

QY 661 CAGTACTCAATCACTGCGGAAATTTTCTGCTGACATTTGATGATTTCTCTGAAAGATATAAG 720
 DB 666 CAGTACTCAATCACTGCGGAAATTTTCTGCTGACATTTGATGATTTCTCTGAAAGATATAAG 725
 QY 721 AGTATCTTTGGACTGGACATCTTTTAAACAGGAGAGAATTTGTGATGTTGCTGGACGAGG 780
 DB 726 AGTATCTTTGGACTGGACATCTTTTAAACAGGAGAGAATTTGTGATGTTGCTGGACGAGG 785
 QY 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
 DB 786 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCTCAGCTGGAATCAGCAA 845
 QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTTAATCAGCTC 900
 DB 846 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTTAATCAGCTC 904
 QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCCTCATCAAT 960
 DB 905 CGACACATCAGCCCTCAAGCC-AAAGCTCTCTTCAGGATAAGGACGTAATTCG-CAWCAAT 962
 QY 961 CAGGACCCCTTGGGCAAGCAAGGG 984
 DB 963 CAGGA-CCCTTGGGCAAGCAAGGK 985

RESULT 9
 BM564282
 LOCUS
 DEFINITION
 5', mRNA sequence.
 BM564282.1 GI:18811955
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo;
 1 (bases 1 to 1067)
 NIH-MGC http://mgc.hci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM12759 row: p column: 07
 High quality sequence stop: 761.
 Location/Qualifiers
 1..1067
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5742198"
 /tissue_type="medulla"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_119"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: EcoRV (destroyed); RNA source normal medulla from
 anonymous male age 27. Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.3 kb, insert size range
 0.9-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 013. Note:
 this is a NIH_MGC Library."

FEATURES
 source

ORIGIN

Query Match		72.3%;	Score 915.8;	DB 3;	Length 1067;
Best Local Similarity		96.4%;	Pred. No. 5.7e-263;		
Matches 1009;		Conservative 0;	Mismatches 31;	Indels 7;	Gaps 7;
QY	6	GCTGAGGACCCAGACATACATCTGGGCTGGCGCTTGGCTTCGGCTTCCTGCGCCCTGGT	65		
DB	15	GCTGAGGACCCAGACATACATCTGGGCTGGCGCTTGGCTTCGGCTTCCTGCGCCCTGGT	74		
QY	66	TTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGGACGCTTACCAT	125		
DB	75	TTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGGACGCTTACCAT	134		
QY	126	GGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACTTGTACCTCCAGGAGAGCCAGATTC	185		
DB	135	GGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACTTGTACCTCCAGGAGAGCCAGATTC	194		
QY	186	CTGCATCAGTGAGAGCTCTTCATGGAGATGCAGAGCTCATGCTCTCAGAGGCTGGA	245		
DB	195	CTGCATCAGTGAGAGCTCTTCATGGAGATGCAGAGCTCATGCTCTCAGAGGCTGGA	254		
QY	246	GGATGAGCTTATGATGACTCTGCAATGATGACTGTTGGATGGCTCCCAAGAGATTC	305		
DB	255	GGATGAGCTTATGATGACTCTGCAATGATGACTGTTGGATGGCTCCCAAGAGATTC	314		
QY	306	AGAAGGACAGCTTCAGGACAGCCCTCAGCGCTTCTCATGGATTCGCCAGCTAGCTAA	365		
DB	315	AGAAGGACAGCTTCAGGACAGCCCTCAGCGCTTCTCATGGATTCGCCAGCTAGCTAA	374		
QY	366	TTATGTTACAGCAAGGAGCTGAAGCTAGGATTTATGAGATGTTGGAAATTAACCTG	425		
DB	375	TTATGTTACAGCAAGGAGCTGAAGCTAGGATTTATGAGATGTTGGAAATTAACCTG	434		
QY	426	CGCAGCTTCCCTGGGAGTTTGGATACTACACATGATGCCAGACCTTTGCTGACTG	485		
DB	435	CGCAGCTTCCCTGGGAGTTTGGATACTACACATGATGCCAGACCTTTGCTGACTG	494		
QY	486	GGGAGTAGATCTGTAATAATTTGATGTTTACTGTGACAGTTTGGAAATTTGGCAGA	545		
DB	495	GGGAGTAGATCTGTAATAATTTGATGTTTACTGTGACAGTTTGGAAATTTGGCAGA	554		
QY	546	TGGTTATAGACATGCTTGGCCCTGATAGGACTGGCAGAGCATGTTGTTACTCTG	605		
DB	555	TGGTTATAGACATGCTTGGCCCTGATAGGACTGGCAGAGCATGTTGTTACTCTG	614		
QY	606	TGAGTGGCTCTTTATATGTGSCCTTTTCAAAGCCCAATATACAGAAATCCGACAGTA	665		
DB	615	TGAGTGGCTCTTTATATGTGSCCTTTTCAAAGCCCAATATACAGAAATCCGACAGTA	674		
QY	666	CTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATAAGAGTAT	725		
DB	675	CTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATAAGAGTAT	734		
QY	726	CTTGACTGGACATCTTTTAAACAGAGAGAAATTTGTTGATGTTGCTGACCA- GGGGTT	784		
DB	735	CTTGACTGGACATCTTTTAA- CAGGAGAGAAATTTGTTGATGTTGCTGACCA- GGGGTT	793		
QY	785	GGAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCTCAGCTGGGAATCAGCAAGTAA	844		
DB	794	GGAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCTCAGCTGGGAATCAGCAAGTAA	853		
QY	845	CTCAGATGGCCCTCTGGGCTATCATGGCTGCT- CCTTTATTCATGCTTAATGACCTCGA	903		
DB	854	CTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCCTTTATTCATGCTTAATGACCTCG	913		
QY	904	CACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGAGCTAATTTGCCATCAATCAG	963		
DB	914	AACATCAACCCCTC- AGCCAAAGCTCTCTTTCAAGATAAGAGCT- ATTGCCATCCATCAG	971		
QY	964	GACCCCTTGGCAAGCAGGTA- CCAGCTTACAGAGGAGCACTTTGAGTGTGGGA	1022		
DB	972	GAAACCTTGGNGCCAGCCAGGGTACCNNCTTAGACCGGAAACAACCTTTTAATGTGGGA	1031		

QY	1023	ACG-ACCTCTCTCAGGCTTAGCCTGGG	1048		
DB	1032	ACGNACTTCTCTCAGGCTTTAACCGGG	1058		
RESULT 10					
AL577581/c		1026 bp	mRNA	linear	EST 07-APR-2004
DEFINITION					
AL577581 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens					
CDNA clone CS0DK009YE12 3-PRIME, mRNA sequence.					
ACCESSION					
AL577581					
VERSION					
AL577581.3 GI:46256603					
KEYWORDS					
EST.					
SOURCE					
Homo sapiens (human)					
ORGANISM					
Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;					
Hominiidae; Homo.					
1 (bases 1 to 1026)					
REFERENCE					
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.					
Full-length cDNA libraries and normalization					
Unpublished (2001)					
On Feb 16, 2001 this sequence version replaced gi:31315840.					
Contact: Genoscope					
Genoscope - Centre National de Sequencage					
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE					
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr					
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime					
end enriched, double-strand cDNA was digested with Not I and cloned					
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library					
was normalized. Library was constructed by Life Technologies, a					
division of Invitrogen. This sequence belongs to sequence cluster					
10506.r					
For more information about this cluster, see					
http://www.genoscope.cns.fr/cdna?s=CS0DK009BC06NP1&c=10506.r.					
FEATURES					
Location/Qualifiers					
1..1026					
/organism="Homo sapiens"					
/mol_type="mRNA"					
/db_xref="taxon:9606"					
/clone="CS0DK009YE12"					
/cell_type="HELA CELLS COT 25-NORMALIZED"					
/cell_lines="HELA"					
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"					
/note="1st strand cDNA was primed with a NotI-oligo(dT)					
primer. Five prime end enriched, double-strand cDNA was					
digested with Not I and cloned into the Not I and EcoR V					
sites of the pCMVSPORT 6 vector. Library was normalized."					
ORIGIN					
Query Match		72.3%;	Score 915;	DB 1;	Length 1026;
Best Local Similarity		98.3%;	Pred. No. 9.8e-263;		
Matches 939;		Conservative 4;	Mismatches 10;	Indels 2;	Gaps 2;
QY	289	GCTCCCCAAGAGATTTCAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCTCATGGG	348		
DB	955	GCTCCCCAAGAGATTTCAGAA-GCAGACTTCA-GCAGACCCCTCAGCGCTTTCTCATGGG	898		
QY	349	ATTGCCAGCTAGCTAATTTATGTTTCACAGCAAGAGACTGAAGCTAGGAGTTTATCAGAT	408		
DB	897	ATTGCCAGCTAGCTAATTTATGTTTCACAGCAAGAGACTGAAGCTAGGAGTTTATCAGAT	838		
QY	409	GTTGAAATAAAACCTGGCAGGCTTCCTCGGAGTTTGGATCTACTACACATTTGATGCC	468		
DB	837	GTTGAAATAAAACCTGGCAGGCTTCCTCGGAGTTTGGATCTACTACACATTTGATGCC	778		
QY	469	CAGACCTTTGCTGACTGGGAGTAGATCTGCTAAATTTGATGTTGTTGTTGTTGTTGTTG	528		
DB	777	CAGACCTTTGCTGACTGGGAGTAGATCTGCTAAATTTGATGTTGTTGTTGTTGTTGTTG	718		
QY	529	TGGAATAATTTGGCAGATGTTTATAGCAGATGCTCTGGCCCTGAATAGGACTGGCAGA	588		
DB	717	TGGAATAATTTGGCAGATGTTTATAGCAGATGCTCTGGCCCTGAATAGGACTGGCAGA	658		

QY	589	AGCATTGTCTACTCTCTGTGAGTGGCCCTCTTTATATGTGSCCCTTTCAAAAGCCCAATTAT	648	AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuina,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsuoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,W., Yoneda,Y., Iihikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
Db	657	ACATTGTGTACTCTCTGTGAGTGGCCCTCTTTATATGTGSCCCTTTCAAAAGCCCAATTAT	598		RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer
QY	649	ACAGAAATCCGACAGTACTGCAATCACTGCGGAAATTTTGTGACATGATGATTCCTGG	708	TITLE	Genome Res. 10 (11), 1757-1771 (2000)
Db	597	ACAGAAATCCGACAGTACTGCAATCACTGCGGAAATTTTGTGACATGATGATTCCTGG	538	JOURNAL	11076861
QY	709	AAAAGTATAAAGAGTATCTTGACATGGACATCTTTTAACGAGAGAGATTTGATGTT	768	PUBMED	
Db	537	AAAAGTATAAAGAGTATCTTGACATGGACATCTTTTAACGAGAGAGATTTGATGTT	478	REFERENCE	
QY	769	GCTGACACAGGGGGTTGGAATGACCCAGATATGTTAGTATGTCGCAACTTTGCGCTCAGC	828	AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Db	477	GCTGACACAGGGGGTTGGAATGACCCAGATATGTTAGTATGTCGCAACTTTGCGCTCAGC	418	TITLE	Functional annotation of a full-length mouse cDNA collection
QY	829	TGGAATCAGCAAGTAACTCAGATGGCCCTCTGCGCTATCATGGCTGCTCTTTATTCATG	888	JOURNAL	Nature 409, 585-590 (2001)
Db	417	TGGAATCAGCAAGTAACTCAGATGGCCCTCTGCGCTATCATGGCTGCTCTTTATTCATG	358	REFERENCE	
QY	889	TCTATGACCTCGACATACAGCTCTCAAGCCAAAGCTCTCTTCAGGATTAAGGACGTA	948	AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Db	357	TCTATGACCTCGACATACAGCTCTCAAGCCAAAGCTCTCTTCAGGATTAAGGACGTA	298	TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
QY	949	ATTGCCATCAATCAGGACCCCTTGGCAAGCAAGGTCACAGCTTAGCAGGAGACAAC	1008	JOURNAL	Nature 420, 563-573 (2002)
Db	297	ATTGCCATCAATCAGGACCCCTTGGCAAGCAAGGTCACAGCTTAGCAGGAGACAAC	238	REFERENCE	6 (bases 1 to 1376)
QY	1009	TTTGAAGTGTGGAAACGACCTCTCTCAGGCTTAGCTGGCTGTAGTATGATAAACCGG	1068	AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayaehida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,Y., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Murakatsu,M. and Hayashizaki,Y.
Db	237	TTTGAAGTGTGGAAACGACCTCTCTCAGGCTTAGCTGGCTGTAGTATGATAAACCGG	178	TITLE	Direct Submission
QY	1069	CAGGAGATTGTGGACCTCGCTCTTATACCATCGCATGCTTCCCTGGGTAAAGAGTG	1128	JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
Db	177	CAGGAGATTGTGGACCTCGCTCTTATACCATCGCATGCTTCCCTGGGTAAAGAGTG	118	COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
QY	1129	GCTGTAATCCTGCTGCTTTCATCACACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTC	1188	FEATURES	source 1. .l376 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:E330039P08" /db_xref="taxon:10090" /clone="E330039P08" /sex="female" /tissue type="ovary" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="2 days pregnant adult"
Db	117	GCTGTAATCCTGCTGCTTTCATCACACAGCTCTCCMTRTGAAAGGAAGCTAGGGTTC	58	CDS	21..1286 /note="unnamed protein product; galactosidase, alpha (MGD MG1:1347344, GB NM_013463, evidence: BLASTN, 99%, match=1373) putative" /codon_start=1 /protein_id="BAC35819.1" /db_xref="GI:26344325"
QY	1189	TATGAATGGACTTCAAGTTAAGAAGTCAATAAATCCACAGGACCTGTTTC	1243		/translation="MAMKLLSRDTRLVCELALCPALVFWLSILGVRLDNGLRPTM GWLHWRFCNLDCQEPDQACISQELFNQMAELMVSQDGRDAGYDLYLCIDCWAPER
Db	57	TATGAATGGACTTCAAGTTAAGAAGTCAATAAATVCRACANNNDTGTGTTTC	3		
RESULT 11	AK054547	1376 bp mRNA linear HTC 03-APR-2004			
LOCUS	AK054547	Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN full-length enriched library, clone:E330039P08			
DEFINITION	AK054547	product:galactosidase, alpha, full insert sequence.			
ACCESSION	AK054547				
VERSION	AK054547.1	GI:26344324			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to Prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
PUBMED	11042159				
REFERENCE	3				

DSKRLQADPPFPSPGSIKHLANYVHSKGLGIYADVGNKTCAGPPGSGFSDIDAQT
 FADWGVDLLKFDGCHDSVSVLENGYKYNALNRTGRSIVYSCSMPLDLPFRPKPNF
 TDIOYCKHWRNDDVDYVSWESIKNLSVYVYQKEIVVEVAGPSWNPDLMLVINGF
 LSWDOQVOTLMALMAAPLANNLNDLROISSOKALLONKQVIAINQDPLGQGYCFR
 KENHLEWVERPLSNLAWAVAVRNLOBIIGPCPYTTIQISSLGRGLACNPGCIITQLLPE
 KYHLGPFYEWTLTKTRVNPSTVLRFLER"
 1358..1363
 /note="putative"
 1376
 /note="putative"

polyA_signal

polyA_site

ORIGIN

Query Match 71.1%; Score 899.8; DB 4; Length 1376;

Best Local Similarity 82.3%; Pred. No. 4.1e-258;

Matches 1033; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTGCGCTTCGCTTCCTGGCC 60
 |||||
 DB 27 ATGAAGCTTTTGAGCAGATACGGCGCTGGCTGTGAGCTTGGCTTTGTCCTTGGCT 86
 |||||
 QY 61 CTCGTTCTCTGGGACATCCCTGGGCTAGAGCAGTGGCAATGGATGGCAAGGACGCT 120
 |||||
 DB 87 TTAGTTTCTGGAGCATTCTGGGCTCAGAGCATTGGCAATGGCTTGGCGGGACTCT 146
 |||||
 QY 121 ACATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGACCA 180
 |||||
 DB 147 ACTATGGGCTGGCTGATTTGGGAACGTTTCATGTGCAACCTTGACTGCCAAGAAGCGCT 206
 |||||
 QY 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 240
 |||||
 DB 207 GATGCTCGATAGTGAGCACTGTTTCATGCAATGGCAGAGCTCATGTCTCTGATGCG 266
 |||||
 QY 241 TGGAGGATGCAGGTTATCAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCCAAGA 300
 |||||
 DB 267 TGGGGGATGCAGTTATGACTATCTCTGCATAGATGACTGTTGGATGGCTCCCGAGG 326
 |||||
 QY 301 GATTGAGAAGGAGCAGCTTCAGGAGACCTTCAGCGCTTCTTCATCGGAGATTCGCCAGCTA 360
 |||||
 DB 327 GATTCAGAGGGCAGGCTTCAGGAGATCCCAACGCTTCTCTAGTGGATCAACACCTC 386
 |||||
 QY 361 GCTAATTATGTCACAGCAAGGACTGAAGCTAGGATTTAGCAGATGTTGGAATAAA 420
 |||||
 DB 387 GCAATATACGTCACAGCAAGGATTTGAAGCTAGGATTTATGCAATGTTGGGAATAA 446
 |||||
 QY 421 ACTGCGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTTGATGCCAGACCTTTGCT 480
 |||||
 DB 447 ACTGTGAGGTTTCCCGGAGTTTGGATCCTATGACATTTGATGGCGAGACATTTGCT 506
 |||||
 QY 481 GACTGGGAGTAGACTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
 |||||
 DB 507 GACTGGGCGTAGACTGCTAAATTTGATGTTGTTGCTGCTGACAGTGTAGTATCTTTG 566
 |||||
 QY 541 GCAGATGTTATAGCACATGCTTGGCCCTGATAGGACTGGCAGAGAGCTTGTGATC 600
 |||||
 DB 567 GAGAATGGTTATAGTACATGGCTTGGCTTGAACAGGACAGGCGGAGAGCATTTGATAC 626
 |||||
 QY 601 TCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGCCCAATATACAGAAATCCGA 660
 |||||
 DB 627 TCTGTGAGTGGCCATTTATTTGAGACCTTTTCATAGCCCAATATACAGATCCAA 686
 |||||
 QY 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGCAATTTGATGATTCCTGGAAAGTATAAG 720
 |||||
 DB 687 TATTACTGCAATCATTTGGAGAAATTTTGTATGATTTTATGATTCCTGGAAAGCATTAAG 746
 |||||
 QY 721 AGTATCTTGACCTGGACATCTTTTAACAGGAGAGAAATTTGTGATGTTGCTGACCGAGG 780
 |||||
 DB 747 AATATCTTGTCTTGGAGAGTGGTTTACAGAGGAGATTTGTTGAAGTCTGCTGGACCGC 806
 |||||
 QY 781 GGTGTGAATGACCCAGATATTTAGTATTTGGCACTTTGGCTCAGCTGGAATCAGCAA 840
 |||||
 DB 807 AGCTGGAATGACCCAGATATTTAGTATTTGGCACTTTGGCTCAGTTGGGACCGAG 866
 |||||
 QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
 |||||

DB 867 GTGACACAGATGGCCCTCTGGGCTATCATGGCCGCTCCCTACTCATGTCCAACGATCTG 926
 |||||
 QY 901 CGACATCAGCCCTCAAGCCAAAAGCTCTCTCTCAGGTAAGGAGCTAATTTGCCATCAAT 960
 |||||
 DB 927 CGCAAAATCAGCTCTCAAGCCAAAAGCTCTCTCTCAGATAAGGATGTAATTTGCCATCAAC 986
 |||||
 QY 961 CAGGACCCCTTGGGCAAGGAGGATCAGCTTAGACAGGAGAGCAACTTTGAAGTGTGG 1020
 |||||
 DB 987 CAAGACCCCTTGGGCAAGCAGGGCTACTGTCTTCAAAAAGGAAAACCATATGAGGTTGG 1046
 |||||
 QY 1021 GAACACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCCGCGAGGAGATTGCT 1080
 |||||
 DB 1047 GAACGCCACCTCTCOACCTAGCTGGGCTGTGGCTGTGAGAAACCTCGAGGAGATTGT 1106
 |||||
 QY 1081 GGACCTCTCTTATATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
 |||||
 DB 1107 GGACCTTGTCTTATACCATCGAGTTTCTTCCCTGGGTAGGAGTACGCTGCAATCTCT 1166
 |||||
 QY 1141 GCCTGCTTCATCAGCAGCTCTCTCTGTGAAAAGGAAAGCTAGGCTTCTATGAATGGACT 1200
 |||||
 DB 1167 GGCTGCATCATTTACTCAGCTTCTCCCGAGAAAGTACACCTAGGCTTCTATGAATGGACT 1226
 |||||
 QY 1201 TCAAGTTAAGAGTACATAAATCCACAGCAGCTGTTTGTCTTCAGCTAGAAA 1255
 |||||
 DB 1227 TTGACCTTAAAACTCGAGTAAACCCCTCAGGCACTGTTTGTTCGGTTAGAAA 1281
 |||||

RESULT 12

AK040081

LOCUS

DEFINITION

AK040081 2962 bp mRNA linear HTC 03-APR-2004
 Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched
 library, clone:A430057F16 product:galactosidase, alpha, full insert
 sequence.

ACCESSION AK040081

VERSION AK040081.1 GI:26333580

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

PUBMED

REFERENCE 10349636

AUTHORS

Carninci, P. and Hayashizaki, Y.
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

PUBMED

REFERENCE 11042159

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

PUBMED

REFERENCE 11076861

AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 695-690 (2001)

JOURNAL

PUBMED

REFERENCE

The FANTOM Consortium and the RIKEN Genome Exploration Research

QY 1201 TCACGGTTAAGTACACATAAATCCACAGGACACTGTTTCTTCAGCTAGAAA 1255
 Db 1207 TTGACCTTAAAAACTCGAGTAACCCCTCAGGCACTGTTTCTTCGGTTAGAAA 1261

RESULT 13

CO645672/c 1133 bp mRNA linear EST 23-JUL-2004
 LOCUS ILLUMIGEN MCQ_25513 Katze MMLV Macaca mulatta cDNA clone
 DEFINITION IBIUW:23923 5' similar to Bases 27 to 1049 highly similar to human
 GLA (Hs.69089), mRNA sequence.

ACCESSION CO645672
 VERSION CO645672.1 GI:50567166
 KEYWORDS EST.
 SOURCE Macaca mulatta (rhesus monkey)

ORGANISM Macaca mulatta
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Cercopithecoidea; Cercopithecinae; Macaca.

REFERENCE 1 (bases 1 to 1133)
 AUTHORS Magness,C.L., Fellin,P.C., Thomas,M.J., Korth,M.J., Agy,M.B.,
 Prohl,S.C., Fitzgibbon,M., Scherer,C.A., Miner,D.G., Katze,M.G. and
 Iadonato,S.P.

TITLE Analysis of the Macaca mulatta transcriptome and the sequence
 divergence between Macaca and human

JOURNAL Genome Biol. 6 (7), R60 (2005)
 PUBMED 15998449
 COMMENT Contact: C. Magness
 Illumigen Biosciences Inc.
 2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
 Tel: 2063780400
 Fax: 2063780408

Email: cmagness@illumigen.com
 Sequenced on 2004.03.05. 699 Q20 bases. Library Preparation: Prof.
 Michael Katze Lab at University of Washington DNA Sequencing:
 Illumigen Biosciences Inc. For further information, see
 http://www.macaque.org

PCR Primers

FORWARD: CCTCACTAAAGGGACAAA
 BACKWARD: CACTATAGGGCGAATTGGTA

Insert Length: 1133 Std Error: 0.00

Plate: CL000111 Row: E Column: 08

Seq primer: CCTCACTAAAGGGACAAA

POLYA=No.

FEATURES

source

Location/Qualifiers

1..1133
 /organism="Macaca mulatta"
 /mol_type="mRNA"
 /strain="Indian"
 /db_xref="taxon:9544"
 /clone="IBIUW:23923"
 /sex="female"
 /dev_stage="adult"
 /lab_host="Electronmax DH10B"
 /clone_lib="katze_MMLV"
 /note="Organ: liver; Vector: pDONR 222; Site 1: BsrG I;
 Site 2: BsrG I; Created from CloneMiner cDNA Library
 Construction kit (catalog #18249-029)"

ORIGIN

Query Match 70.3%; Score 890; DB 7; Length 1133;
 Best Local Similarity 94.3%; Pred. No. 3.3e-255;
 Matches 945; Conservative 0; Mismatches 55; Indels 2; Gaps 2;
 QY 264 CCTCTGCATTGATGACTGTTGGATGCTCCCAAGAGATTTCAGAGGACGACTTCAGGC 323
 Db 1041 CCTCTGCCCCCGGACTTTTGGAGATTCCCAAGAGACTCGAACGGCCCTTCGGGC 982
 QY 324 AGACCTCTAGCGCTTTCCTCATGGGATTCGCCAGCTAGCTAATTATGTTTCACAGCAAGG 383
 Db 981 GGGCGGTAGTGTCTTCTTCATGGGATTCGCCAGTAGCTAATTATGATACAGCAAGG 922
 QY 384 ACTGAAGCTAGGAGTTTATGACAGATGTTGGAAATAAAACCTCGCAGGCTTCCTCCGGAG 443

Db 921 AATGAAATAGGGATTTATGACAGCGTTGGAAATAAAACCTCGCAGGCTTCCTTGGGAG 862
 QY 444 TTTTGGATACACACATTGATGCCACACCTTTGCTGACTGGGAGTAGATCTCTATAA 503
 Db 861 TTTTGGATACATGACATTGATGCCACCTCTTGTCTGCTGAGTAGATATGTTGAA 802
 QY 504 ATTTGATGGTTGTTACTGTGACAGTTTGGAAAAATTTGGCAGATGTTTATAGCACATGTC 563
 Db 801 ATTTGATGGTTGTTACTGTGACAGTTTGGAAAAATTTGGCAGATGTTTATAGCACATGTC 742
 QY 564 CTGGCCCTGAATAGGACTGGCAGAGCATTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 623
 Db 741 GTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 682
 QY 624 GTGGCCCTTTCAAAAGCCCAATTATPACAGAAATCCGACAGTACTGCAATCACTGGCGAAA 683
 Db 681 GTGGCCATTTCAAAAGCCCAATTATPACAGAAATCCGACAGTACTGCAATCACTGGCGAAA 622
 QY 684 TTTTGTGACATTTGATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 743
 Db 621 TTTTGTGACATTTGATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 562
 QY 744 TAACCCAGGAGAGATTGTTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 803
 Db 561 TAACCCAGGAGAGATTGTTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 502
 QY 804 AGTGATTTGGCAATTTTGGCCCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 863
 Db 501 AGTGATTTGGCAATTTTGGCCCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 442
 QY 864 TATCATGCTGCTCCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 923
 Db 441 TATCATGCTGCTCCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 382
 QY 924 AGCTCTCCTTCAGGATAGGACGTAATTTGCCATCAATCAGGACCCCTTGGGCAAGCAAGG 983
 Db 381 AGCTCTCCTTCAGGATAGGACGTAATTTGCCATCAATCAGGACCCCTTGGGCAAGCAAGG 322
 QY 984 GTACAGCTTTAGACAGGGAGAGACAACTTTTGAAGTGTGGAAACGACCTCTCTCAGGCTTAGC 1043
 Db 321 GTACAGCTTTAGACAGGGAGAGACAACTTTTGAAGTGTGGAAACGACCTCTCTCAGGCTTAGC 262
 QY 1044 CTGGGCTGTAGCTATGATAAACCAGCAGAGATTTGGTGGACCTCGCTCTTTATACCATCGC 1103
 Db 261 CTGGGCTGTAGCTATGATAAACCAGCAGAGATTTGGTGGACCTCGCTCTTTATACCATCGC 202
 QY 1104 AGTTGCTTCCCTGGTAAAGGAGTGGCTGTAATCTCGCTGCTTCATCACACAGCTCCT 1163
 Db 201 GGTGCTTCCCTGGTAAAGGAGTGGCTGTAATCTCGCTGCTTCATCACACAGCTCCT 142
 QY 1164 CCCTGTGAAAAGGAGCTAGGCTTCTATGAATGGAATTTCAAGGTTTAAAGATCACATAAA 1223
 Db 141 CCCTGTGAAAAGGAGCTAGGCTTCTATGAATGGAATTTCAAGGTTTAAAGATCACATAAA 83
 QY 1224 TCCACAGGCACTGTTTGTCTTTCAGCTAGAGTAAACCAATGTA 1265
 Db 82 TCCACAGGCACTG-TTTGCTTGGCTAGAAAAATACAATGCA 42

RESULT 14
 CO645623
 LOCUS ILLUMIGEN MCQ_24642 Katze MMLV Macaca mulatta cDNA clone
 DEFINITION IBIUW:23923 5' similar to Bases 5 to 920 highly similar to human
 GLA (Hs.69089), mRNA sequence.
 ACCESSION CO645623
 VERSION CO645623.1 GI:50567117
 KEYWORDS EST.
 SOURCE Macaca mulatta (rhesus monkey)
 ORGANISM Macaca mulatta
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

/clone="CS0DI067YJ01"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 69.6%; Score 880.8; DB 1; Length 1071;
Best Local Similarity 91.6%; Pred. No. 1.9e-252;
Matches 985; Conservative 20; Mismatches 61; Indels 9; Gaps 8;

QY	191	TCAGTGAGAGGCTCTTCATGAGATGGCAGAGCTCATGCTCTCAGAA-GGCTGGAAGGAT	249
DB	1071	TCAGTGAGAGGCTC-TCATGNANATGNCAGANCTCATGNTCTCAGAACGMCTGAAAGNAT	1013
QY	250	GCAGGTTATGATACCTCTGC--ATTGATGACTGTTGGATGGCT-CCCCAAGAGATTCA	306
DB	1012	GCAGTNTATGATACCTCTGCANTTTGATGACTGTTGNATGNTCTCCCCAAAGAGATNCA	953
QY	307	GAAAGCAGACTTTCAGCAGACCTCAGCGCTTTTCCTCATGGGATTCGCCAGCTAGCTAAT	366
DB	952	GAAAGCAGACTTTCAGCAGACCTCAGCGCTTTTCCTCATGGGATTCGCCAGCTAGCTAAT	893
QY	367	TATGTTTCACAGCAAGGACTGAAGCTAGGGATTTATGACAGATGTTGGAAATAAAACCTGC	426
DB	892	TATGTTTCACAGCAAGGACTGAAGCTA-GGATTTATGACAGATGTT-GAAATAAAACCTGC	835
QY	427	GCAGGCTCCCTGGGAGTTTGGATACGACTACGACATTCATGATGATGATGATGATGATGAT	486
DB	834	GCAGGCTCCCTGGGAGTTTGGATACGACTACGACATTCATGATGATGATGATGATGATGAT	775
QY	487	GCAGTAGATCTGCTAAATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	546
DB	774	GCAGTAGATCTGCTAAATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	715
QY	547	GGTTATAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTTACTCCTGT	606
DB	714	GATMATAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTTACTCCTGT	655
QY	607	GAGTGGCTCTTTATATGATGGCCCTTTTCAAAAGCCCAAT-TATACAGAAATCCGACAGTA	665
DB	654	VAGTRACCTCACTATATCTGCCCCCTTMYCAAAAGMCCAATMTATACAAAAATCCGACAGTA	595
QY	666	CTGCAATCACTGGGAAATTTGCTGACATTCATGATTCCTGCAAAAGTATAAAGAGTAT	725
DB	594	CTGCAATCACTGGGAAATTTGCTGACATTCATGATTCCTGCAAAAGTATAAAGAGTAT	535
QY	726	CTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACCAGGGGGTTG	785
DB	534	CTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACCAGGGGGTTG	475
QY	786	GAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAAGTAAC	845
DB	474	GAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAASTMAC	415
QY	846	TCAGATGGCCCTCTGGGCTCATGCTGCTCTCTTTTATTCATGCTCTAATGACCTCCGACA	905
DB	414	TCAGAT-GGCCCTCTGGGCTCATGCTGCTCTCTTTTATTCATGCTCTAATGACCTCCGACA	356
QY	906	CATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAGGACGTAATTTGCCATCAATCAGGA	965
DB	355	CATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAGGACGTAATTTGCCATCAATCAGGA	296
QY	966	CCCTTTGGGCAAGCAAGGGTACAGCTTAGACAGGAGACAACCTTTGAAAGTGTGGGAACG	1025
DB	295	CCCTTTGGGCAAGCAAGGGTACCRGCTTAGACAGGAGACAACCTTTGAAAGTGTGGGAACG	236
QY	1026	ACCTCTCTCAGGCTTAGGCTGTAGCTATGATATAAACCGCAGGAGATTTGGTGACC	1085
DB	235	ACVTCTCTCAGGCTTAGGCTGTAGCTATGATATAAACCGCAGGAGATTTGGTGACC	176

Search completed: December 26, 2005, 08:29:58
Job time : 4823 secs

QY	1086	TCGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCCTGTATCTCTGCTG	1145
DB	175	TCGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCCTGTATCTCTGCTG	116
QY	1146	CTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACTTCAAG	1205
DB	115	CTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACTTCAAG	56
QY	1206	GTTAAGAAAGTCAATATAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAAACACA	1260
DB	55	GTTAAGAAAGTCAATATAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAAATACA	1

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 04:55:19 ; Search time 264 Seconds
(without alignments)
8524.216 Million cell updates/sec

Title: US-10-602-219-11
Perfect score: 1266
Sequence: 1 atcgagtgtgaggaaccacaga.....agctagaacacaaatgtaa 1266

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCUS COMB.seq.*
7: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	100.0	1266	3	US-09-993-059-11
2	1266	100.0	1266	3	US-10-103-327-11
3	1264	99.8	1284	3	US-09-993-059-13
4	1264	99.8	1284	3	US-10-103-327-13
5	1263.4	99.8	1278	3	US-09-993-059-7
6	1263.4	99.8	1278	3	US-10-103-327-7
7	1263.4	99.8	1290	3	US-09-993-059-3
8	1263.4	99.8	1290	3	US-10-103-327-3
9	1263.4	99.8	1296	3	US-09-993-059-9
10	1263.4	99.8	1296	3	US-10-103-327-9
11	1263.4	99.8	1308	3	US-09-993-059-5
12	1263.4	99.8	1308	3	US-10-103-327-5
13	1261.8	99.7	1343	3	US-08-928-881-18
14	1261.8	99.7	1343	3	US-09-543-921-18
15	1261.8	99.7	1343	3	US-09-266-014-3
16	1261.8	99.7	1343	3	US-09-491-759-18
17	1261.8	99.7	1393	2	US-07-602-824A-1
18	1261.8	99.7	1393	2	US-07-983-451-1
19	1261.8	99.7	1393	2	US-08-261-577-6
20	1252.4	98.9	1254	3	US-09-993-059-15
21	1252.4	98.9	1254	3	US-10-103-327-15
22	1251	98.8	1272	3	US-09-993-059-17
23	1251	98.8	1272	3	US-10-103-327-17
24	1251	98.8	8234	3	US-09-626-127-14

ALIGNMENTS

RESULT 1

US-09-993-059-11
; Sequence 11, Application US/09993059
; Patent No. 6887896
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUNAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-09-993-059-11

Query Match	100.0%	Score 1266;	DB 3;	Length 1266;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1266;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
Qy	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGCTGCGGCTTGGCGCTTGGCTTCTTGGCC	60	
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGCTGCGGCTTGGCGCTTGGCTTCTTGGCC	60	
Qy	61	CTCGTTTCTGGGACATCCCTGGGCTAGACACTGGCAATGGATTGGCAGGACGCT	120	
Db	61	CTCGTTTCTGGGACATCCCTGGGCTAGACACTGGCAATGGATTGGCAGGACGCT	120	
Qy	121	ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA	180	
Db	121	ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA	180	
Qy	181	GATTCCTGCATCAGTGAGAACTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240	
Db	181	GATTCCTGCATCAGTGAGAACTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240	
Qy	241	TGGAGGATGAGGTTATGACTCTCTGCATTGATGACTGTGGATGGCTCCCAAGA	300	
Db	241	TGGAGGATGAGGTTATGACTCTCTGCATTGATGACTGTGGATGGCTCCCAAGA	300	

```
QY 301 GATTTCAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATATATGTTTCACAGCAAGAGACTGAAGCTAGGGAATTAATGACAGATGTTGGAATAAA 420
Db 361 GCTAATATATGTTTCACAGCAAGAGACTGAAGCTAGGGAATTAATGACAGATGTTGGAATAAA 420
QY 421 ACTGCGGAGGCTTCCTGGGAGTTTGTGATACCTAGCAATGATGAGCCAGACCTTTGCT 480
Db 421 ACTGCGGAGGCTTCCTGGGAGTTTGTGATACCTAGCAATGATGAGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGAGCAGATTTGGAATAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGAGCAGATTTGGAATAATTG 540
QY 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTATC 600
Db 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTATC 600
QY 601 TCCTGTAGTGGCTCTTTATATGATGAGCCCTTTCAAAGCCCAATTAATACAGAAATCCGA 660
Db 601 TCCTGTAGTGGCTCTTTATATGATGAGCCCTTTCAAAGCCCAATTAATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAATTTGCTGACATTTGATGATTCCTGGAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAATTTGCTGACATTTGATGATTCCTGGAAAGTATAAG 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATGAGTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTATGAGTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTTCGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTTCGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTAACAGCTTAGACAGGAGACAACCTTTGAAGTGTG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTAACAGCTTAGACAGGAGACAACCTTTGAAGTGTG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATATAACCGGAGAGATTTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATATAACCGGAGAGATTTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
QY 1141 GCCTGTTCATCACAGCTCTCCTGTGTAAGGAGAGAGAGTGGGTTCTATGAATGGACT 1200
Db 1141 GCCTGTTCATCACAGCTCTCCTGTGTAAGGAGAGAGTGGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCACATAATCCACAGGACATGTTTCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCACATAATCCACAGGACATGTTTCTTCAGCTAGAAAACACA 1260
QY 1261 ATGTAA 1266
Db 1261 ATGTAA 1266
```

RESULT 2

US-10-103-327-11

; Sequence 11, Application US/10103327

```
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 08010087CUP06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-103-327-11
```

```
Query Match 100.0%; Score 1266; DB 3; Length 1266;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTTCCTGCCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTTCCTGCCC 60
QY 61 CTCGTTTCTGGACATCCCTGGGCTAGACACTGGCAATGATTTGGCAAGACGCT 120
Db 61 CTCGTTTCTGGACATCCCTGGGCTAGACACTGGCAATGATTTGGCAAGACGCT 120
QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCA 180
QY 181 GATTCCTGTCATCAGTGAAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCCTGTCATCAGTGAAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGGAAGGATGAGGTATGAGTACCTCTGCAATTTGATGATGCTGTTGGATGGCTCCCAAGA 300
Db 241 TGGAAGGATGAGGTATGAGTACCTCTGCAATTTGATGATGCTGTTGGATGGCTCCCAAGA 300
QY 301 GATTCAGAAAGGAGACTTCAGGAGAGCCCTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAGAAAGGAGACTTCAGGAGAGCCCTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATATGTTTCACAGCAAGAGACTGAAGCTAGGGAATTAATGACAGATGTTGGAATAAA 420
Db 361 GCTAATATGTTTCACAGCAAGAGACTGAAGCTAGGGAATTAATGACAGATGTTGGAATAAA 420
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTAGGACATGATGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTAGGACATGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTGTGTTACTGTGACAGATTTGGAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTGTGTTACTGTGACAGATTTGGAAATTTG 540
QY 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTATC 600
Db 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTATC 600
QY 601 TCCTGTAGTGGCTCTTTATATGATGAGCCCTTTCAAAGCCCAATTAATACAGAAATCCGA 660
Db 601 TCCTGTAGTGGCTCTTTATATGATGAGCCCTTTCAAAGCCCAATTAATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAATTTGCTGACATTTGATGATTCCTGGAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAATTTGCTGACATTTGATGATTCCTGGAAAGTATAAG 720
```

```
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAGATATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATTTGTCTGGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATTTGTCTGGACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTAGTATTTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTAGTATTTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTGCTCTTTATTCATGCTCTAAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTGCTCTTTATTCATGCTCTAAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGTAAATGGCCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGTAAATGGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAAAGCACTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
Db 1021 GAAAGCACTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
Qy 1141 GCCTGCTTCAACACAGCTCTCCTCTGTGAAAAGAGTGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCAACACAGCTCTCCTCTGTGAAAAGAGTGGCTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCAATAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260
Qy 1261 ATGTAA 1266
Db 1261 ATGTAA 1266
```

RESULT 3

```
US-09-993-059-13
; Sequence 13, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
; US-09-993-059-13
```

```
Query Match 99.8%; Score 1264; DB 3; Length 1284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTCGGCC 60
|||||
```

```
Db 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTCGGCC 60
Qy 61 CTCGTTTCTGGGACATCCCTTGGGGCTAGAGCACTGGACAATGGATTGGCAAGCAGCCT 120
|||||
Db 61 CTCGTTTCTGGGACATCCCTTGGGGCTAGAGCACTGGACAATGGATTGGCAAGCAGCCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCGCAGGAAGAGCA 180
|||||
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCGCAGGAAGAGCA 180
Qy 181 GATTCTCTGCATCAGTGTGAGAGCTCTTCATGAGAGATGGCAGAGCTCATGGTCTCAGNAGGC 240
|||||
Db 181 GATTCTCTGCATCAGTGTGAGAGCTCTTCATGAGAGATGGCAGAGCTCATGGTCTCAGNAGGC 240
Qy 241 TGGGAAGGATGAGGTTATGAGTACCTCTGCAATTTGATGACTGTTGGATGGCTCCCAAGA 300
|||||
Db 241 TGGGAAGGATGAGGTTATGAGTACCTCTGCAATTTGATGACTGTTGGATGGCTCCCAAGA 300
Qy 301 GATTTCAGAGGAGCAGCTTTCAGGAGAGCCCTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
|||||
Db 301 GATTTCAGAGGAGCAGCTTTCAGGAGAGCCCTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTTATGTTTACAGCAAGAGGACTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
|||||
Db 361 GCTAATTTATGTTTACAGCAAGAGGACTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATACACGACATTTGATGCCAGACCTTTGCT 480
|||||
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATACACGACATTTGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGGATGATCTGCTAATAATTTGATGGTTTACTGTGACAGTTTGGAAATTTG 540
|||||
Db 481 GACTGGGAGGATGATCTGCTAATAATTTGATGGTTTACTGTGACAGTTTGGAAATTTG 540
Qy 541 GCAGATGTTTATAGCACATGCTTGGCCCTGTAATAGGACTGCGCAGAGCAATTTGTGTAC 600
|||||
Db 541 GCAGATGTTTATAGCACATGCTTGGCCCTGTAATAGGACTGCGCAGAGCAATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCCCTCTTTATATATGTCGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
|||||
Db 601 TCCTGTGAGTGGCCCTCTTTATATATGTCGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGCGGAAATTTTGTGCTGACATTTGATGATTTCTGGAAGTATAAAG 720
|||||
Db 661 CAGTACTGCAATCACTGCGGAAATTTTGTGCTGACATTTGATGATTTCTGGAAGTATAAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGTCTGGACAGGG 780
|||||
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGTCTGGACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTAGTATTTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
|||||
Db 781 GGTGGAAATGACCCAGATATGTTAGTAGTATTTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTCTAAATGACCTC 900
|||||
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTCTAAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGTAAATGGCCATCAAT 960
|||||
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGTAAATGGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
|||||
Db 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAAAGCACTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
|||||
Db 1021 GAAAGCACTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
|||||
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
```



```
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-09-993-059-7

Query Match      99.8%; Score 1263.4; DB 3; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTGCTTCTGCGC 60
DB 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTGCTTCTGCGC 60
QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGAACAATGGATTGGCAAGAGCGCT 120
DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGAACAATGGATTGGCAAGAGCGCT 120
QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
DB 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
QY 181 GATTCTGTCATCAGTGAAGAGCTTTCATGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
DB 181 GATTCTGTCATCAGTGAAGAGCTTTCATGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
QY 241 TGGAGGATGAGGATTTATGAGTACCTCTGCAATGTATGATGACTGTGGATGGCTCCCCAAGA 300
DB 241 TGGAGGATGAGGATTTATGAGTACCTCTGCAATGTATGATGACTGTGGATGGCTCCCCAAGA 300
QY 301 GATTGAGGAGAGACTTTCAGGAGAGCCCTGAGCGCTTTCATCATGGAATTCGCCAGCTA 360
DB 301 GATTGAGGAGAGACTTTCAGGAGAGCCCTGAGCGCTTTCATCATGGAATTCGCCAGCTA 360
QY 361 GCTAATATGTTTACAGCAAGAGGAGCTGAAGCTTAGGATTTATGCAAGATGTTGGAAATAA 420
DB 361 GCTAATATGTTTACAGCAAGAGGAGCTGAAGCTTAGGATTTATGCAAGATGTTGGAAATAA 420
QY 421 ACCTGCGGAGGCTTCCCTGGGAGTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
DB 421 ACCTGCGGAGGCTTCCCTGGGAGTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGATTTGGAAATTTG 540
DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGATTTGGAAATTTG 540
QY 541 GCAGATGTTTAAAGCAGATGCTTGGCCCTGGAATAGGAGCTGGCAGAGCATTTGTGTAC 600
DB 541 GCAGATGTTTAAAGCAGATGCTTGGCCCTGGAATAGGAGCTGGCAGAGCATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTTGGGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTTGGGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGGCAAAATTTGCTGACATTTGATGATTCCTGGAAAGATATAAG 720
DB 661 CAGTACTGCAATCACTGGGCAAAATTTGCTGACATTTGATGATTCCTGGAAAGATATAAG 720
QY 721 AGTATCTTGGCTGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
DB 721 AGTATCTTGGCTGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTCGAATTTGGCCTCAGCTGGAATCAGCAA 840
DB 781 GGTGGAATGACCCAGATATGTTAGTATGTCGAATTTGGCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACATGAGTGGCCTCTGGGCTATCATGGCTGCTCTCTTTATTTATGCTTAATGACCTC 900
DB 841 GTAACATGAGTGGCCTCTGGGCTATCATGGCTGCTCTCTTTATTTATGCTTAATGACCTC 900
```

```
DB 841 GTAACATGAGTGGCCTCTGGGCTATCATGGCTGCTCTCTTTATTTATGCTTAATGACCTC 900
QY 901 CGACATCATGAGCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTTGCCATCAAT 960
DB 901 CGACATCATGAGCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGGAGGTACAGCTTAGACAGGAGACAACTTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGGAGGTACAGCTTAGACAGGAGACAACTTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGGACGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGGACGAGATTGGT 1080
QY 1081 GGACCTCGCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCCTGTAATCCT 1140
DB 1081 GGACCTCGCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCCTGTAATCCT 1140
QY 1141 GCCTGCTTATCACACAGCTCTCTCTGTGTAAGGAGAGCTAGGTTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTATCACACAGCTCTCTCTGTGTAAGGAGAGCTAGGTTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTTAAAGAGTCAATAAATCCACAGGCACTGTTTTGCTTACGCTAGAAAAACACA 1260
DB 1201 TCAAGGTTTAAAGAGTCAATAAATCCACAGGCACTGTTTTGCTTACGCTAGAAAAACACA 1260
QY 1261 ATGTA 1265
DB 1261 ATGCA 1265
```

RESULT 6

```
US-10-103-327-7
; Sequence 7, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-10-103-327-7
```

```
Query Match      99.8%; Score 1263.4; DB 3; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTGCTTCTGCGC 60
DB 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTGCTTCTGCGC 60
QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGAACAATGGATTGGCAGGAGCGCT 120
DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGAACAATGGATTGGCAGGAGCGCT 120
QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
DB 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
```


Db 601 TCCTGTGAGTGGCCCTCTTTATATGTGGCCCTTTTCAAAAGCCCAATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGGAAATTTTCTGACATTTGATGATTTCTGCGAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGGAAATTTTCTGACATTTGATGATTTCTGCGAAAGTATAAG 720
Qy 721 AGTATCTTGGACATGACATCTTTTAAACAGAGAGAAATTTGTGATTTGCTGGACAGGG 780
Db 721 AGTATCTTGGACATGACATCTTTTAAACAGAGAGAAATTTGTGATTTGCTGGACAGGG 780
Qy 781 GGTGGATGACCCAGATGATTTAGTGATTTGCACTTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGATGACCCAGATGATTTAGTGATTTGCACTTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGTAAATGGCCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGTAAATGGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAGTGTGG 1020
Qy 1021 GAAACACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAACCAGGAGAGATGGT 1080
Db 1021 GAAACACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAACCAGGAGAGATGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAGGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAGGAGTGGCTGTAATCCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCTCTGCTGTAAGGAGAGCTAGGTTCTTGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCTCTGCTGTAAGGAGAGCTAGGTTCTTGAATGGACT 1200
Qy 1201 TCAAGTTTAAAGATGACATAAATCCACAGGCTGTTTCTTCACTAGAAAACACA 1260
Db 1201 TCAAGTTTAAAGATGACATAAATCCACAGGCTGTTTCTTCACTAGAAAACACA 1260
Qy 1261 ATGTA 1265
Db 1261 ATGTA 1265

RESULT 8
US-10-103-327-3
; Sequence 3, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-10-103-327-3

Query Match 99.8%; Score 1263.4; DB 3; Length 1290;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTCTGCCC 60
Db 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTCTGCCC 60
Qy 61 CTGCTTTCTGGGACATCCCTCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGACGCT 120
Db 61 CTGCTTTCTGGGACATCCCTCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGACGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGGCTTCTATGTGCAACCTTGTGCTGCGAAGAGCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGGCTTCTATGTGCAACCTTGTGCTGCGAAGAGCA 180
Qy 181 GATTCTGTCATCAGTGAGAGCTCTTCTATGAGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Db 181 GATTCTGTCATCAGTGAGAGCTCTTCTATGAGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Qy 241 TGGAGGATGAGGCTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300
Db 241 TGGAGGATGAGGCTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300
Qy 301 GATTCAAGGACGACTTTCAGGAGAGCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAAGGACGACTTTCAGGAGAGCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTTATGTTTACAGCAAGGACTGAGCTAGGATTTATGCAAGTGTGGAAATAAA 420
Db 361 GCTAATTTATGTTTACAGCAAGGACTGAGCTAGGATTTATGCAAGTGTGGAAATAAA 420
Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACACGACATTTGATGCGCCAGACCTTGT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACACGACATTTGATGCGCCAGACCTTGT 480
Qy 481 GACTGGGAGTGTGCTTAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAATTTG 540
Db 481 GACTGGGAGTGTGCTTAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 541 GCAGATGGTTATAGCACATGCTTGGCCCTGTAATAGGACTGGCAGAGCAATCTGTGAC 600
Db 541 GCAGATGGTTATAGCACATGCTTGGCCCTGTAATAGGACTGGCAGAGCAATCTGTGAC 600
Qy 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAAAGCCCAATATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAAAGCCCAATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGGAAATTTTGTGCTGACATTTGATGATTTCTGCGAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGGAAATTTTGTGCTGACATTTGATGATTTCTGCGAAAGTATAAG 720
Qy 721 AGTATCTTGGACATGACATCTTTTAAACAGAGAGAAATTTGTGATTTGCTGGACAGGG 780
Db 721 AGTATCTTGGACATGACATCTTTTAAACAGAGAGAAATTTGTGATTTGCTGGACAGGG 780
Qy 781 GGTGGATGACCCAGATGATTTAGTGATTTGCACTTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGATGACCCAGATGATTTAGTGATTTGCACTTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGTAAATGGCCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGTAAATGGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAGTGTGG 1020

```
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTCCTCCCTGGGTAAGAGAGTGGCCTGTAAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGAGTTCCTCCCTGGGTAAGAGAGTGGCCTGTAAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAAGGAGCTAGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAAGGAGCTAGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
QY 1261 ATGTA 1265
Db 1261 ATGCA 1265

RESULT 9
US-09-993-059-9
; Sequence 9, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1296
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1296)
US-09-993-059-9

Query Match 99.8%; Score 1263.4; DB 3; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTGGCC 60
QY 61 CTCGTTCTTCGGGACATCCTCGGGCTAGAGCACTGGACAATGATTTGGCAAGGACGCT 120
Db 61 CTCGTTCTTCGGGACATCCTCGGGCTAGAGCACTGGACAATGATTTGGCAAGGACGCT 120
QY 121 ACATGGGCTGGTGACCTGGGAGGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
Db 121 ACCATGGGCTGGTGACCTGGGAGGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGTGGCAGAGCTCATGGTCTCAGNAGGC 240
Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGTGGCAGAGCTCATGGTCTCAGNAGGC 240
QY 241 TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
QY 301 GATTGAGAGGACGACTTCAGGAGACCTTCAGCGCTTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTGAGAGGACGACTTCAGGAGACCTTCAGCGCTTTTCCTCATGGGATTCGCCAGCTA 360
```

```
QY 361 GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGAGATGTTTGGAAATAAA 420
Db 361 GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGAGATGTTTGGAAATAAA 420
QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATPACTACGACATTTGATGCTCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATPACTACGACATTTGATGCTCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTTACTGTGACAGTTTGGMAAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTTACTGTGACAGTTTGGMAAATTG 540
QY 541 GCAGATGTTTATAAGCACATGCTCCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600
Db 541 GCAGATGTTTATAAGCACATGCTCCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCCTCTTTTATATGTTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCCTCTTTTATATGTTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCTGACATTTGATGATTCCTGGAAGATATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCTGACATTTGATGATTCCTGGAAGATATAAG 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGCAGAGAGATTGTTGATGTTTGTGTCGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGCAGAGAGATTGTTGATGTTTGTGTCGACAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTATTTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATTTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCCTTTTATTCATGCTCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCCTTTTATTCATGCTCTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCCTCTTCAAGGATAAGGAGCTAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCCTCTTCAAGGATAAGGAGCTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTTGGGCAAGGGGTACAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTTGGGCAAGGGGTACAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAAAGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
Db 1021 GAAAGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCCTGTAAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCCTGTAAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAAGGAGCTAGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAAGGAGCTAGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
QY 1261 ATGTA 1265
Db 1261 ATGCA 1265
```

```
RESULT 10
US-10-103-327-9
; Sequence 9, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
```

;; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
;; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
;; FILE REFERENCE: 008010087CPUS06
;; CURRENT APPLICATION NUMBER: US/10/103.327
;; CURRENT FILING DATE: 2002-03-20
;; PRIOR APPLICATION NUMBER: US/09/993.059
;; PRIOR FILING DATE: 2001-11-13
;; NUMBER OF SEQ ID NOS: 37
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 9
;; LENGTH: 1296
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(1296)
US-10-103-327-9

Query Match 99.8%; Score 1263.4; DB 3; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGGCTTGGCTTCTGGCC 60
Db 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGGCTTGGCTTCTGGCC 60

Qy 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGAATGGATTGGCAAGCGCT 120
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGAATGGATTGGCAAGCGCT 120

Qy 121 ACCATGGCTGGCTGACCTGGAGGCTTCTATGCAACCTTGAATGGATTGGCAAGCGCT 180
Db 121 ACCATGGCTGGCTGACCTGGAGGCTTCTATGCAACCTTGAATGGATTGGCAAGCGCT 180

Qy 181 GATTCCTGCATCAGTGAAGCTTCTCATGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
Db 181 GATTCCTGCATCAGTGAAGCTTCTCATGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240

Qy 241 TGGAGGATGAGGTTATGAGTACCTCTGCATTTGATGATGCTTGGATGGCTCCCAAGA 300
Db 241 TGGAGGATGAGGTTATGAGTACCTCTGCATTTGATGATGCTTGGATGGCTCCCAAGA 300

Qy 301 GATTCAGAGGAGAGCTTTCAGGAGACCTCTGAGGCTTCTCATGGATTGGCCAGCTA 360
Db 301 GATTCAGAGGAGAGCTTTCAGGAGACCTCTGAGGCTTCTCATGGATTGGCCAGCTA 360

Qy 361 GCTAATTTATGTTACAGCAAGGAGCTGAAGCTAGGATTTATGCAAGATTTGGAAATAA 420
Db 361 GCTAATTTATGTTACAGCAAGGAGCTGAAGCTAGGATTTATGCAAGATTTGGAAATAA 420

Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACGACATTTGATGCTCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACGACATTTGATGCTCCAGACCTTTGCT 480

Qy 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTATCTGACAGATTTGGAAAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTATCTGACAGATTTGGAAAATTG 540

Qy 541 GCAGATGTTATAGCAGATCTCTGCTCCCTGATAGGCTGGCAGAGCATTTGTGTAC 600
Db 541 GCAGATGTTATAGCAGATCTCTGCTCCCTGATAGGCTGGCAGAGCATTTGTGTAC 600

Qy 601 TCCTGTAGTGGCTCTTTATATGTTGGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTAGTGGCTCTTTATATGTTGGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660

Qy 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTTCTGAAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTTCTGAAAAGTATAAG 720

Qy 721 AGTATCTGGACTGACATCTTTTAAACAGGAGAGATTTGATGTTGCTGGACACGG 780
Db 721 AGTATCTGGACTGACATCTTTTAAACAGGAGAGATTTGATGTTGCTGGACACGG 780

Qy 781 GGTGGAAATGACCCAGATATGTTAGTGAATTTGGCAACTTTGGCCTCAGCTGGAATCAGCA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATTTGGCAACTTTGGCCTCAGCTGGAATCAGCA 840

Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTCTAATGACCTC 900

Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCCAGGATAAGAGCTAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCCAGGATAAGAGCTAATTTGCCATCAAT 960

Qy 961 CAGGACCCCTTGGGCAAGAGGTAACAGCTTATAGACAGGAGACAATTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGAGGTAACAGCTTATAGACAGGAGACAATTTGAAGTGTGG 1020

Qy 1021 GAACGACCTCTCTCAGGCTTACCTGGGCTAGTATGATAAAACCGCAGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTACCTGGGCTAGTATGATAAAACCGCAGAGATTGGT 1080

Qy 1081 GGACCTCGCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGGAGTGGCTTGAATCCT 1140
Db 1081 GGACCTCGCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGGAGTGGCTTGAATCCT 1140

Qy 1141 GCCTGCTTATCAGACAGCTCTCCCTGTAAGAGGAGTGGCTTGAATGAGTGGT 1200
Db 1141 GCCTGCTTATCAGACAGCTCTCCCTGTAAGAGGAGTGGCTTGAATGAGTGGT 1200

Qy 1201 TCAAGGTTAAGAGTCAATAAATCCCAAGGAGTGGTGGCTTGGCTTGAATGAGTGGT 1260
Db 1201 TCAAGGTTAAGAGTCAATAAATCCCAAGGAGTGGTGGCTTGGCTTGAATGAGTGGT 1260

Qy 1261 ATGTA 1265
Db 1261 ATGCA 1265

RESULT 11

US-09-993-059-5
; Sequence 5, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993.059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1308
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1308)
US-09-993-059-5

Query Match 99.8%; Score 1263.4; DB 3; Length 1308;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGCCTTGGCTTCTGGCC 60
Db 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGGCTTGGCCTTGGCTTCTGGCC 60

Qy 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGAATGGATTGGCAAGGAGCGCT 120
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGAATGGATTGGCAAGGAGCGCT 120

121 ACCATGGCTGGCTGACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180
Db |||||
121 ACCATGGCTGGCTGACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180
Qy |||||
181 GATTCTCGCATCAGTGAAGGCTCTTCATGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db |||||
181 GATTCTCGCATCAGTGAAGGCTCTTCATGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy |||||
241 TGAAGGATGAGGTTATGAGTACCTCTGCAATGATGACTGTGTGATGGCTCCCAAGA 300
Db |||||
241 TGAAGGATGAGGTTATGAGTACCTCTGCAATGATGACTGTGTGATGGCTCCCAAGA 300
Qy |||||
301 GATTGAGGAGGAGCTTCAGGACACCTCAGCGCTTCCTCATGGGATGGCCAGCTA 360
Db |||||
301 GATTGAGGAGGAGCTTCAGGACACCTCAGCGCTTCCTCATGGGATGGCCAGCTA 360
Qy |||||
361 GCTAATTATGTTACAGCAAGGACTGAAGCTAGGATTTATGCAAGTGTGGAATAAA 420
Db |||||
361 GCTAATTATGTTACAGCAAGGACTGAAGCTAGGATTTATGCAAGTGTGGAATAAA 420
Qy |||||
421 ACTCGGAGGCTTCCTGGAGTTTGGATCTACGACATGATGCCAGACCTTTGCT 480
Db |||||
421 ACTCGGAGGCTTCCTGGAGTTTGGATCTACGACATGATGCCAGACCTTTGCT 480
Qy |||||
481 GACTGGGAGTAGATCTGTAATAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTG 540
Db |||||
481 GACTGGGAGTAGATCTGTAATAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTG 540
Qy |||||
541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGAC 600
Db |||||
541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGAC 600
Qy |||||
601 TCCTGTAGTGGCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTATACAGAAATCGA 660
Db |||||
601 TCCTGTAGTGGCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTATACAGAAATCGA 660
Qy |||||
661 CAGTACTGAATCACTGGCGAATTTGCTGACATTTGATGATCTGCTGGAAGTAAAG 720
Db |||||
661 CAGTACTGAATCACTGGCGAATTTGCTGACATTTGATGATCTGCTGGAAGTAAAG 720
Qy |||||
721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTTGATGTTGCTGACAGGG 780
Db |||||
721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTTGATGTTGCTGACAGGG 780
Qy |||||
781 GGTGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
Db |||||
781 GGTGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
Qy |||||
841 GTAACTCAGATGGCCCTCTGGGCTATCAGCTGCTCCTTTTATTCATGCTTAATGACCTC 900
Db |||||
841 GTAACTCAGATGGCCCTCTGGGCTATCAGCTGCTCCTTTTATTCATGCTTAATGACCTC 900
Qy |||||
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGAGCTAATTCGCATCAAT 960
Db |||||
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGAGCTAATTCGCATCAAT 960
Qy |||||
961 CAGGACCCCTTGGGCAAGCAAGGTTACGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db |||||
961 CAGGACCCCTTGGGCAAGCAAGGTTACGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy |||||
1021 GAACGACCTCTCTCAGGCTTAGCGGCTGTAGCTATGATATAACCGGCAGAGATTGGT 1080
Db |||||
1021 GAACGACCTCTCTCAGGCTTAGCGGCTGTAGCTATGATATAACCGGCAGAGATTGGT 1080
Qy |||||
1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTTGGGTAAGGAGTGGCCCTGTAATCCT 1140
Db |||||
1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTTGGGTAAGGAGTGGCCCTGTAATCCT 1140
Qy |||||
1141 GCGTGTTCATACAGCTCCTCCCTGTGAAAGGAACTAGGTTCTATGAATGGACT 1200
Db |||||
1141 GCGTGTTCATACAGCTCCTCCCTGTGAAAGGAACTAGGTTCTATGAATGGACT 1200

Qy 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCCTTCACTAGAAAACACA 1260
Db |||||
1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCCTTCACTAGAAAACACA 1260
Qy 1261 ATGTA 1265
Db |||||
1261 ATGCA 1265

RESULT 12
US-10-103-327-5
; Sequence 5, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUNAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1308
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1308)
US-10-103-327-5

Query Match 99.8%; Score 1263.4; DB 3; Length 1308;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
Db |||||
1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
Qy 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGAATGGCAAGGACGCT 120
Db |||||
61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGAATGGCAAGGACGCT 120
Qy 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGAATGCCAGGAAGGCA 180
Db |||||
121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGAATGCCAGGAAGGCA 180
Qy 181 GATTCTCGATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db |||||
181 GATTCTCGATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 241 TGAAGGATGAGGTTATGAGTACCTCTGCAATGATGACTGTGTGATGGCTCCCAAGA 300
Db |||||
241 TGAAGGATGAGGTTATGAGTACCTCTGCAATGATGACTGTGTGATGGCTCCCAAGA 300
Qy 301 GATTGAGGAGGAGCTTCAGGACACCTCAGCGCTTCCTCATGGGATTCGCCAGCTA 360
Db |||||
301 GATTGAGGAGGAGCTTCAGGACACCTCAGCGCTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTATGTTACAGCAAGGACTGAAGCTAGGATTTATGCAAGTGTGGAATAAA 420
Db |||||
361 GCTAATTATGTTACAGCAAGGACTGAAGCTAGGATTTATGCAAGTGTGGAATAAA 420
Qy 421 ACCTCGGAGGCTTCCTCGGAGTTTGGATCTACGACATGATGCCAGACCTTTGCT 480
Db |||||
421 ACCTCGGAGGCTTCCTCGGAGTTTGGATCTACGACATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGCTAATAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTG 540
Db |||||

QY 721 ACTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTCGACCGG 780
Db 756 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTCGACCGG 815
QY 781 GCTTGGAAATGACCCAGATATGTTAGTGATTTGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
Db 816 GCTTGGAAATGACCCAGATATGTTAGTGATTTGGCACTTTGGCCTCAGCTGGAATCAGCAA 875
QY 841 GTAACCTAGATGGCCTCTGGGCTATCATGCTGCTGCTCTTTTATCATGCTTAATGACCTC 900
Db 876 GTAACCTAGATGGCCTCTGGGCTATCATGCTGCTGCTCTTTTATCATGCTTAATGACCTC 935
QY 901 CCACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGTAATTCGCATCAAT 960
Db 936 CCACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGTAATTCGCATCAAT 995
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTTAGACAGGGAGACAACTTTGAACTGTGG 1020
Db 996 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTTAGACAGGGAGACAACTTTGAACTGTGG 1055
QY 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGAGCTATGATTAATCAACCGGAGAGATTTGCT 1080
Db 1056 GAACGACCTCTCTCAGGCTTAGCTGGGCTGAGCTATGATTAATCAACCGGAGAGATTTGCT 1115
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
Db 1116 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1175
QY 1141 GCCTGCTTCATCACAGCTCCTCCTCTGTAAGGAGAGTGGCTGTAACTGGAATGGAAT 1200
Db 1176 GCCTGCTTCATCACAGCTCCTCCTCTGTAAGGAGAGTGGCTGTAACTGGAATGGAAT 1235
QY 1201 TCAAGGTTAAGAGTCACATAATCCCAAGGACCTGTTTCTCAGCTAGAGAAACACA 1260
Db 1236 TCAAGGTTAAGAGTCACATAATCCCAAGGACCTGTTTCTCAGCTAGAGAAACACA 1295
QY 1261 ATGTA 1265
Db 1296 ATGCA 1300

RESULT 14
US-09-543-921-18
; Sequence 18, Application US/09543921
; Patent No. 6395884
; GENERAL INFORMATION:
; APPLICANT: Selden et al., Richard F.
; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A
; DEFICIENCY
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/543,921
; FILING DATE: 06-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,881
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819

; REFERENCE/DOCKET NUMBER: 07236/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1343 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-543-921-18

Query Match 99.7%; Score 1261.8; DB 3; Length 1343;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1263; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGCAAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTGGCCTTTCGCTTCTCGGCC 60
Db 36 ATGCAAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTGGCCTTTCGCTTCTCGGCC 95
QY 61 CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGAACAATGGATTTGGCAAGAGCGCT 120
Db 96 CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGAACAATGGATTTGGCAAGAGCGCT 155
QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCACTGTGCAACCTTGACTGCCAGGAAGAGCA 180
Db 156 ACCATGGGCTGGCTGCACTGGGAGCGCTTCACTGTGCAACCTTGACTGCCAGGAAGAGCA 215
QY 181 GATTCTCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 216 GATTCTCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 275
QY 241 TGGAAAGGATCAGGTTTATGATGATCCTCTGCAATTTGATGATCTGTGATGGCTCCCCAAGA 300
Db 276 TGGAAAGGATCAGGTTTATGATGATCCTCTGCAATTTGATGATCTGTGATGGCTCCCCAAGA 335
QY 301 GATTTCAGAGGAGAGACTTTCAGGAGCGCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 336 GATTTCAGAGGAGAGACTTTCAGGAGCGCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 395
QY 361 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGATTTATGCAAGATGTTGGAATATAA 420
Db 396 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGATTTATGCAAGATGTTGGAATATAA 455
QY 421 ACCTCGGAGGCTTCCCTGGGAGTTTGGATACATGACATTTGATGATGATGATGATGATGAT 480
Db 456 ACCTCGGAGGCTTCCCTGGGAGTTTGGATACATGACATTTGATGATGATGATGATGATGAT 515
QY 481 GACTGGGAGTAGATCTGTAATAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 540
Db 516 GACTGGGAGTAGATCTGTAATAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 575
QY 541 GCAGATGGTTATAGCACATGTCCTTGGCCCTGAAATAGGACTGGCAGAGCAATTTGTGTAC 600
Db 576 GCAGATGGTTATAGCACATGTCCTTGGCCCTGAAATAGGACTGGCAGAGCAATTTGTGTAC 635
QY 601 TCCTGTGATGGCTCTTTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
Db 636 TCCTGTGATGGCTCTTTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 695
QY 661 CAGTACTGCAATCACTGGGAGAAATTTTCTGCACTTGTGATGATGATGATGATGATGATGATGAT 720
Db 696 CAGTACTGCAATCACTGGGAGAAATTTTCTGCACTTGTGATGATGATGATGATGATGATGATGAT 755
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATGATGATGATGATGATGAT 780
Db 756 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATGATGATGATGATGATGAT 815
QY 781 GGTGGAATGACCCAGATATGTTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Db 816 GGTGGAATGACCCAGATATGTTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 875

QY 841 GTAACATGAGTGGCCCTCTGGGCTATCATGCTGCTCCTTTATTCATGCTCTAATGACCTC 900
DB 876 GTAACATGAGTGGCCCTCTGGGCTATCATGCTGCTCCTTTATTCATGCTCTAATGACCTC 935
QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAAGGATAAGGAGCTAATTTGCCATCAAT 960
DB 936 CGACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAAGGATAAGGAGCTAATTTGCCATCAAT 995
QY 961 CAGGACCCCTTGGGCAAGCAAGGTPACAGCTTACAGAGGAGACAACTTTGAAAGTGTGG 1020
DB 996 CAGGACCCCTTGGGCAAGCAAGGTPACAGCTTACAGAGGAGACAACTTTGAAAGTGTGG 1055
QY 1021 GAAAGACCTCTCTCAGGCTTACCTGGGCTAGCTATGATATAAACCAGGAGAGATTGGT 1080
DB 1056 GAAAGACCTCTCTCAGGCTTACCTGGGCTAGCTATGATATAAACCAGGAGAGATTGGT 1115
QY 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
DB 1116 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1175
QY 1141 GCCTGCTTATCACACAGCTCCTCCTCTGTGAAAAGGAGCTAGGGTCTTATGAATGGACT 1200
DB 1176 GCCTGCTTATCACACAGCTCCTCCTCTGTGAAAAGGAGCTAGGGTCTTATGAATGGACT 1235
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGGCTTACGCTAGAAAAACACA 1260
DB 1236 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGGCTTACGCTAGAAAAATACA 1295
QY 1261 ATGTA 1265
DB 1296 ATGCA 1300

RESULT 15

US-09-266-014-3

; Sequence 3, Application US/09266014

; Patent No. 6458574

; GENERAL INFORMATION:

; APPLICANT: Selden, Richard F

; APPLICANT: Borowski, Marianne

; APPLICANT: Kinoshita, Carol M

; APPLICANT: Treco, Douglas A

; APPLICANT: Williams, Melanie D

; APPLICANT: Schuetz, Thomas J

; APPLICANT: Daniel, Peter F.

; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency

; FILE REFERENCE: FABRY DISEASE (18082-001)

; CURRENT APPLICATION NUMBER: US/09/266,014

; PRIORITY FILING DATE: 1999-03-11

; PRIORITY APPLICATION NUMBER: 60/026,041

; PRIORITY FILING DATE: 1996-09-13

; PRIORITY APPLICATION NUMBER: 08/928,881

; PRIORITY FILING DATE: 1997-09-12

; PRIORITY APPLICATION NUMBER: PCT US97/16603

; PRIORITY FILING DATE: 1997-09-12

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 1343

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-266-014-3

Query Match 99.7%; Score 1261.8; DB 3; Length 1343;

Best Local Similarity 99.8%; Pred. No. 0; Mismatches 2; Indels 0; Gaps 0;
Matches 1263; Conservative 0;

QY 1 ATGAGCTGAGGAACCCAGAACTACATCTGGGCTCGGGCTGCGCTTTCGCTTCCTGGCC 60

DB 36 ATGAGCTGAGGAACCCAGAACTACATCTGGGCTCGGGCTTTCGCTTCCTGGCC 95

QY 61 CTCGTTCTTGGGACATCCCTGGGCTAGAGCACTGGAGTATGGCAAGGAGCGCT 120

|||||

DB 96 CTCGTTCTTGGGACATCCCTGGGCTAGAGCACTGGCAATGATTTGGCAAGGAGCGCT 155
QY 121 ACCATGGGCTGGTGCACCTGGGAGCGCTTCATGTGCAACCTTTGACTGCGCAGGAAGAGCA 180
DB 156 ACCATGGGCTGGTGCACCTGGGAGCGCTTCATGTGCAACCTTTGACTGCGCAGGAAGAGCA 215
QY 181 GATTCTGTCATCAGTGGAGAGCTCTTCATGAGAGATGGCAGAGCTCATGGTCTCAGGAAGGC 240
DB 216 GATTCTGTCATCAGTGGAGAGCTCTTCATGAGAGATGGCAGAGCTCATGGTCTCAGGAAGGC 275
QY 241 TGGGAAGATGAGGCTTATGATGACTCTGCAATTCATGACTCTTGGATGGCTCCCAAGA 300
DB 276 TGGGAAGATGAGGCTTATGATGACTCTGCAATTCATGACTCTTGGATGGCTCCCAAGA 335
QY 301 GATTCTGAGGAGCACTTCAGGAGAGCCCTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB 336 GATTCTGAGGAGCACTTCAGGAGAGCCCTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 395
QY 361 GCTAATTTATGTTTACAGCAAGGACTGAAGCTAGGGATTTATGAGATGTTTGGAAATAA 420
DB 396 GCTAATTTATGTTTACAGCAAGGACTGAAGCTAGGGATTTATGAGATGTTTGGAAATAA 455
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTCCTGAGTACTACGACATTCATGCTGCGCAGCTTTCCT 480
DB 456 ACCTGGCAGGCTTCCCTGGGAGTTCCTGAGTACTACGACATTCATGCTGCGCAGCTTTCCT 515
QY 481 GACTGGGAGTATGCTGCTAATAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAATTTG 540
DB 516 GACTGGGAGTATGCTGCTAATAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAATTTG 575
QY 541 GCAGATGGTTATAGCAGATGCTCTTGGCCCTGATAGGACTGGCAGAGCACTTGTGTAC 600
DB 576 GCAGATGGTTATAGCAGATGCTCTTGGCCCTGATAGGACTGGCAGAGCACTTGTGTAC 635
QY 601 TCCTGTGAGTGGCTCTTTTATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB 636 TCCTGTGAGTGGCTCTTTTATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 695
QY 661 CAGTACTGCAATCACTGCGGAAATTTTGTGCTGACATTCATGATTCCTGAAAAGTATAAG 720
DB 696 CAGTACTGCAATCACTGCGGAAATTTTGTGCTGACATTCATGATTCCTGAAAAGTATAAG 755
QY 721 AGTATCTTGGATCGGACATCTTTTAAACAGAGAGATTTGTTGTTGCTGGACAGGG 780
DB 756 AGTATCTTGGATCGGACATCTTTTAAACAGAGAGATTTGTTGTTGCTGGACAGGG 815
QY 781 GTTTGGAAATGACCCAGATATGTTAGTATGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
DB 816 GTTTGGAAATGACCCAGATATGTTAGTATGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCA 875
QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCCTTTTATTCATGCTCTAATGACCTC 900
DB 876 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCCTTTTATTCATGCTCTAATGACCTC 935
QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGCTAATTTGCCATCAAT 960
DB 936 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGCTAATTTGCCATCAAT 995
QY 961 CAGGACCCCTTGGGCAAGCAAGGTPACAGCTTACAGAGGAGACAACTTTGAAAGTGTGG 1020
DB 996 CAGGACCCCTTGGGCAAGCAAGGTPACAGCTTACAGAGGAGACAACTTTGAAAGTGTGG 1055
QY 1021 GAAAGACCTCTCTCAGGCTTACCTGGGCTAGCTATGATATAAACCAGGAGAGATTGGT 1080
DB 1056 GAAAGACCTCTCTCAGGCTTACCTGGGCTAGCTATGATATAAACCAGGAGAGATTGGT 1115
QY 1081 GGAAGCTGCTTATACCATCGAGTTCCTTCCCTGGGTAAGGAGTGGCTTCTGTAATCCT 1140
DB 1116 GGAAGCTGCTTATACCATCGAGTTCCTTCCCTGGGTAAGGAGTGGCTTCTGTAATCCT 1175
QY 1141 GCCTGCTTATCACACAGCTCCTCCTGTGAAAAGGAGCTAGGGTCTTATGAATGGACT 1200
DB 1176 GCCTGCTTATCACACAGCTCCTCCTGTGAAAAGGAGCTAGGGTCTTATGAATGGACT 1235

Qy 1201 TCAAGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTAGAAAACACA 1260
Db 1236 TCAAGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTAGAAAATACA 1295

Qy 1261 ATGTA 1265
Db 1296 ATGCA 1300

Search completed: December 26, 2005, 08:34:32
Job time : 268 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 07:09:39 ; Search time 1004 Seconds
(without alignments)
10427.323 Million cell updates/sec

Title: US-10-602-219-11
Perfect score: 1266
Sequence: 1 atgcagctgaggaaccagca.....agctagaaacacaaatgtaa 1266

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 9793542 seqs, 4134689005 residues
Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA_Main:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	100.0	1266	3	US-09-993-059-11
2	1266	100.0	1266	5	US-10-103-327-11
3	1266	100.0	1266	7	US-10-602-219-11
4	1266	100.0	1266	8	US-10-602-220-11
5	1266	100.0	1266	9	US-10-984-389-11
6	1266	100.0	1266	8	US-10-984-388-11
7	1264	99.8	1284	5	US-09-993-059-13
8	1264	99.8	1284	5	US-10-103-327-13
9	1264	99.8	1284	7	US-10-602-219-13
10	1264	99.8	1284	8	US-10-602-220-13
11	1264	99.8	1284	8	US-10-851-388-13
12	1264	99.8	1284	9	US-10-984-389-13
13	1263.4	99.8	1278	3	US-09-993-059-7
14	1263.4	99.8	1278	5	US-10-103-327-7
15	1263.4	99.8	1278	7	US-10-602-219-7
16	1263.4	99.8	1278	7	US-10-602-220-7
17	1263.4	99.8	1278	8	US-10-851-388-7
18	1263.4	99.8	1278	9	US-10-984-389-7
19	1263.4	99.8	1290	5	US-09-993-059-3
20	1263.4	99.8	1290	5	US-10-103-327-3
21	1263.4	99.8	1290	7	US-10-602-219-3
22	1263.4	99.8	1290	7	US-10-602-220-3
23	1263.4	99.8	1290	8	US-10-851-388-3

ALIGNMENTS

RESULT 1
US-09-993-059-11
; Sequence 11, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-09-993-059-11

Query Match	100.0%	Score 1266;	DB 3;	Length 1266;
Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches 0;	Indels 0;
Matches 1266;	Conservative	0;	0;	0;
Qy	1	ATGCAGCTGAGGAACCCAGAACACTACATCTGGGCTGGCGCTTGGCGCTTCGCTTCCTGGCC	60	
Db	1	ATGCAGCTGAGGAACCCAGAACACTACATCTGGGCTGGCGCTTGGCGCTTCGCTTCCTGGCC	60	
Qy	61	CTCGTTTCTGGGACATCCCTGGGGCTAGACACTGGACATGGATTGGCAAGNCGCCT	120	
Db	61	CTCGTTTCTGGGACATCCCTGGGGCTAGACACTGGACATGGATTGGCAAGNCGCCT	120	
Qy	121	ACCATGGGCTGGCTGGAGCGCTTTCATGTGCAACCTTGACTGCTCCAGGAGGCA	180	
Db	121	ACCATGGGCTGGCTGGAGCGCTTTCATGTGCAACCTTGACTGCTCCAGGAGGCA	180	
Qy	181	GATTCTTCGATCAGTGAGAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240	
Db	181	GATTCTTCGATCAGTGAGAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240	
Qy	241	TGGAAGGATGAGGTTATGAGTACTCTTCGATTGATGATCTTGGATGGCTCCCAAGA	300	
Db	241	TGGAAGGATGAGGTTATGAGTACTCTTCGATTGATGATCTTGGATGGCTCCCAAGA	300	

```
Db 241 TGAAGGATGACAGTTATGAGTACCTCTGCAATTGATGACTGTTGATGGCTCCCAAAGA 300
Qy 301 GATTTCAGAGGAGACTTCAGGACAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAGGAGAGACTTCAGGACAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTATGTTTCACAGCAAGAGACTGAAGCTAGGAGTTTATGTCAGATGTTTGGAAATAAA 420
Db 361 GCTAATTATGTTTCACAGCAAGAGACTGAAGCTAGGAGTTTATGTCAGATGTTTGGAAATAAA 420
Qy 421 ACCTCGCAGGCTTCCCTGGGAGTTTGGATACAGCAATGATGCCAGACCTTTGCT 480
Db 421 ACCTCGCAGGCTTCCCTGGGAGTTTGGATACAGCAATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGATTTGGAAAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGATTTGGAAAATTG 540
Qy 541 GCAGATGGTTATAGACACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600
Db 541 GCAGATGGTTATAGACACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTTATATGTCGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTTATATGTCGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCAGCTGGCGAAATTTGCTGCACATGATGATTCCTGGAAAGTATAAG 720
Db 661 CAGTACTGCAATCAGCTGGCGAAATTTGCTGCACATGATGATTCCTGGAAAGTATAAG 720
Qy 721 AGTATCTTGGACTGACATCTTTTAAACAGAGAGAGAAATTTGTTGATGTTGCTGGACACGG 780
Db 721 AGTATCTTGGACTGACATCTTTTAAACAGAGAGAGAAATTTGTTGATGTTGCTGGACACGG 780
Qy 781 GGTTCGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Db 781 GGTTCGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATCATGTCATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATCATGTCATGACCTC 900
Qy 901 GCACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGCTAATTCGCCATCAAT 960
Db 901 GCACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGAGCTAATTCGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTTAGACAGGGAGACAACCTTTGAAAGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTTAGACAGGGAGACAACCTTTGAAAGTGG 1020
Qy 1021 GAACGACCTCTCAGGCTTAGCCTTAGCCTGGCTGTAGCTATGATATAACCGGCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCAGGCTTAGCCTTAGCCTGGCTGTAGCTATGATATAACCGGCAGGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCCTGTAATCCT 1140
Qy 1141 GCCTGCTTCAACACAGCTCCTCCTGTGAAAAGGAGCTAGGGTTCCTATGAATGGACT 1200
Db 1141 GCCTGCTTCAACACAGCTCCTCCTGTGAAAAGGAGCTAGGGTTCCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAGATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCAGATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
Qy 1261 ATGTAA 1266
Db 1261 ATGTAA 1266
```

```
; Sequence 11, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CUP06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-103-327-11
```

```
Query Match 100.0%; Score 1266; DB 5; Length 1266;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTGGCGCTTTCCTGCTTCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTTCCTGCTTCTGGCC 60

Qy 61 CTCGTTTCTTGGEACATCCCTGGGCTAGAGCACTGGCAATGGATTTGGCAAGGACGCT 120
Db 61 CTCGTTTCTTGGEACATCCCTGGGCTAGAGCACTGGCAATGGATTTGGCAAGGACGCT 120

Qy 121 ACCATGGGCTGGCTGCACTGGGAGGCTTCATGTGCAACCTTGTACTGCCAGGAAGAGCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGGCTTCATGTGCAACCTTGTACTGCCAGGAAGAGCA 180

Qy 181 GATTCTCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Db 181 GATTCTCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240

Qy 241 TGGAAAGATGAGGTTATGATGACTCTGCAATGATGATGATGATGATGATGATGATGATG 300
Db 241 TGGAAAGATGAGGTTATGATGACTCTGCAATGATGATGATGATGATGATGATGATGATG 300

Qy 301 GATTTCAGAAAGCAGACTTCAGGACAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAAAGCAGACTTCAGGACAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

Qy 361 GCTAATTATGTTTCACAGCAAGAGACTGAAGCTAGGAGTTTATGTCAGATGTTTGGAAATAAA 420
Db 361 GCTAATTATGTTTCACAGCAAGAGACTGAAGCTAGGAGTTTATGTCAGATGTTTGGAAATAAA 420

Qy 421 ACCTCGCAGGCTTCCCTGGGAGTTTGGATACAGCAATGATGCCAGACCTTTGCT 480
Db 421 ACCTCGCAGGCTTCCCTGGGAGTTTGGATACAGCAATGATGCCAGACCTTTGCT 480

Qy 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGATTTGGAAAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGATTTGGAAAATTG 540

Qy 541 GCAGATGGTTATAGACACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600
Db 541 GCAGATGGTTATAGACACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTCGTAC 600

Qy 601 TCCTGTGAGTGGCTCTTTTATATGTCGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTTATATGTCGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660

Qy 661 CAGTACTGCAATCAGCTGGCGAAATTTGCTGCACATGATGATTCCTGGAAAGTATAAG 720
```

661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATTTGATGATTCCTGGAAAGTATAAG 720
721 AGTATCTTGGACTGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACACGG 780
721 AGTATCTTGGACTGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACACGG 780
781 GGTGGAAATGACCCAGATATGTTAGTATGTTAGTATGTTGGCTCAGCTGGAATCAGAA 840
781 GGTGGAAATGACCCAGATATGTTAGTATGTTAGTATGTTGGCTCAGCTGGAATCAGAA 840
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTATTCATCTCTAAATGACCTC 900
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTATTCATCTCTAAATGACCTC 900
901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGCTAATGGCCATCAAT 960
901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGAGCTAATGGCCATCAAT 960
961 CAGGACCCCTGGGCAAGAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
961 CAGGACCCCTGGGCAAGAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
1021 GAACGACCTCTCTAGGCTTAGCTGGCTGTAGCTATGATAAACCAGGAGATTTGGT 1080
1021 GAACGACCTCTCTAGGCTTAGCTGGCTGTAGCTATGATAAACCAGGAGATTTGGT 1080
1081 GGACCTCGCTCTTATACCATGCACTGCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
1081 GGACCTCGCTCTTATACCATGCACTGCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGGAGTGGCTGTATGAAATGGACT 1200
1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGGAGTGGCTGTATGAAATGGACT 1200
1201 TCAGGTTAAGAGTACATTAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260
1201 TCAGGTTAAGAGTACATTAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260
1261 ATGTAA 1266
1261 ATGTAA 1266

RESULT 3
US-10-602-219-11
; Sequence 11, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692

; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-11

Query Match 100.0%; Score 1266; DB 7; Length 1266;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTTGGCC 60
Qy 61 CTCGTTTCTCGGACATCCCTGGGCTAGACCTAGACCTAGGCAATGGATTTGCAAGGACGCT 120
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGACCTAGGCAATGGATTTGCAAGGACGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGGCTTTCATGTGCAACCTTTGACTGCCAGGAAGCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGGCTTTCATGTGCAACCTTTGACTGCCAGGAAGCA 180
Qy 181 GATTCTGTCATCAGTGAGAAAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCTGTCATCAGTGAGAAAGCTCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATTTGATGACTGTTGATGCTGCTCCCAAGA 300
Db 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATTTGATGACTGTTGATGCTGCTCCCAAGA 300
Qy 301 GATTCAAGAGCAGACTTTCAGGAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAAGAGCAGACTTTCAGGAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTTATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
Db 361 GCTAATTTATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
Qy 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACGACATTTGATGCTGCAAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACGACATTTGATGCTGCAAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Db 481 GACTGGGAGTAGATCTCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Qy 541 GCAGATGTTTATAGCACATGCTCTGGCCCTGAAATAGGACTGGCAGAAAGCAATTTGTAC 600
Db 541 GCAGATGTTTATAGCACATGCTCTGGCCCTGAAATAGGACTGGCAGAAAGCAATTTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTTATATGCTGGCCCTTCAAAAGCCCAATATATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTTATATGCTGGCCCTTCAAAAGCCCAATATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTGGAAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTGGAAAAGTATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGTTAGTATGTTGGCTCAGCTGGAATCAGCA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTTAGTATGTTGGCTCAGCTGGAATCAGCA 840

```
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCCTTTATTTATTCATGCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCCTTTATTTATTCATGCTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTAGGATAAGGAGTAAATGGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTAGGATAAGGAGTAAATGGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGAGGTACAGCTTTAGACAGGGAGACAACTTTGAAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGAGGTACAGCTTTAGACAGGGAGACAACTTTGAAAGTGTGG 1020
QY 1021 GAAACACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCGCAGGAGATGGT 1080
Db 1021 GAAACACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCGCAGGAGATGGT 1080
QY 1081 GGACCTCGCTTTATACCATCGAGTTGCTCCCTGGGTAAAGGAGTGGCCTGTAACTCT 1140
Db 1081 GGACCTCGCTTTATACCATCGAGTTGCTCCCTGGGTAAAGGAGTGGCCTGTAACTCT 1140
QY 1141 GCTCGCTTCATCACACAGCTCTCCCTGTGTAAGAGGAGTGGGTTCTATGAATGGACT 1200
Db 1141 GCTCGCTTCATCACACAGCTCTCCCTGTGTAAGAGGAGTGGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCACATAATCCACAGGCACTGTTTCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAGTCACATAATCCACAGGCACTGTTTCTTCAGCTAGAAAACACA 1260
QY 1261 ATGTAA 1266
Db 1261 ATGTAA 1266
```

RESULT 4

```
US-10-602-220-11
; Sequence 11, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT FILING DATE: 2003-06-23
; PRIOR FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
```

```
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-220-11
```

```
Query Match 100.0%; Score 1266; DB 7; Length 1266;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCCTGCTGCCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCCTGCTGCCC 60
QY 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACATGGACAATGGATTTGGCAAGGACGCT 120
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACATGGACAATGGATTTGGCAAGGACGCT 120
QY 121 ACCATGGGCTGGCTGCTGCTGAGGAGGCTTTCATGTGCAACCTTTGACTGCCAGGAAGCCA 180
Db 121 ACCATGGGCTGGCTGCTGCTGAGGAGGCTTTCATGTGCAACCTTTGACTGCCAGGAAGCCA 180
QY 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATTTGATGACTGTTGGATGCTCCCAAGA 300
Db 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATTTGATGACTGTTGGATGCTCCCAAGA 300
QY 301 GATTGAGGAGGAGGAGCTTTCAGGAGAGCTTTCAGGAGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTGAGGAGGAGGAGCTTTCAGGAGAGCTTTCAGGAGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTATGTTACAGCAAGAGGAGCTGAAAGCTAGGAGTTTATGCAAGATGTTGGAAATAA 420
Db 361 GCTAATTATGTTACAGCAAGAGGAGCTGAAAGCTAGGAGTTTATGCAAGATGTTGGAAATAA 420
QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTCGATACAGCATTTGATGCTCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTCGATACAGCATTTGATGCTCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGATTTGGAATAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGATTTGGAATAATTG 540
QY 541 GCAGATGTTTATAGCACATGCTCTTGGCCCTGTAATAGCATGCGCAGAGCAATTTGTATC 600
Db 541 GCAGATGTTTATAGCACATGCTCTTGGCCCTGTAATAGCATGCGCAGAGCAATTTGTATC 600
QY 601 TCCTGTGAGTGGCCCTTTTATATGTTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCCCTTTTATATGTTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGCGGAAATTTTGTGACATTTGATGTTCTGGAATAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGCGGAAATTTTGTGACATTTGATGTTCTGGAATAAGTATAAG 720
QY 721 AGTATCTTGGAGTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db 721 AGTATCTTGGAGTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTGGCCTCAGCTGGAAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTGGCCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCCTTTATTTATTCATGCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCCTTTATTTATTCATGCTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTAGGATAAGGAGTAAATGGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTAGGATAAGGAGTAAATGGCCATCAAT 960
```

QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGAGCAACTTTGAAGTGGG 1020
Db |
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGAGCAACTTTGAAGTGGG 1020
Db |
QY 1021 GAAAGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
Db |
QY 1021 GAAAGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
Db |
QY 1081 GGACCTCGCTCTTATPACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Db |
QY 1081 GGACCTCGCTCTTATPACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Db |
QY 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAGGAGTGGCTGTAAATGAGT 1200
Db |
QY 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAGGAGTGGCTGTAAATGAGT 1200
Db |
QY 1201 TCAAGGTTAAGAGTACATAAATCCCAAGGAGTGGCTGTAAATGAGT 1260
Db |
QY 1261 ATGTAA 1266
Db |
QY 1261 ATGTAA 1266
Db |

RESULT 5

US-10-851-388-11

; Sequence 11, Application US/10851388

; Publication No. US20040234516A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/851,388

; CURRENT FILING DATE: 2004-05-21

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11

; LENGTH: 1266

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(1266)

US-10-851-388-11

Query Match 100.0%; Score 1266; DB 8; Length 1266;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGACGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTGGCTTCTCGGCC 60
Db |
QY 1 ATGACGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTGGCTTCTCGGCC 60
Db |
QY 61 CTCGTTTCTGGGACATCGCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCGCT 120
Db |
QY 61 CTCGTTTCTGGGACATCGCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCGCT 120
Db |
QY 121 ACCATGGGCTGCTGCTGAGGAGCGCTTCATGTGCAACCTTGGCTGAGGAGGAGCA 180
Db |
QY 121 ACCATGGGCTGCTGCTGAGGAGCGCTTCATGTGCAACCTTGGCTGAGGAGGAGCA 180
Db |
QY 181 GATTCCTGCTCAGTGAGAGCTCTTATGAGAGTGGCAGAGCTCATGTTCTCAGAGGC 240
Db |
QY 181 GATTCCTGCTCAGTGAGAGCTCTTATGAGAGTGGCAGAGCTCATGTTCTCAGAGGC 240
Db |
QY 241 TGGAGAGATGCAAGTTATGAGTACCTCTGCAATGATGATGTTGATGGCTCCCAAGA 300
Db |

RESULT 6

US-10-984-389-11


```
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGGCTTCGCTTCCTGGCC 60
Qy 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGATGGCAAGGAGCGCT 120
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGATGGCAAGGAGCGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180
Db 121 ACCATGGGCTGGCTGCACTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180
Qy 181 GATTCTCTGATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db 181 GATTCTCTGATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Qy 301 GATTGAGAGGAGCACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTGAGAGGAGCACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTATGTTACAGCAAGGACTGAGCTAGGATTTATGCAAGATTTGGAATATAA 420
Db 361 GCTAATTATGTTACAGCAAGGACTGAGCTAGGATTTATGCAAGATTTGGAATATAA 420
Qy 421 ACCTCGGAGGCTTCCTCGGAGTTTGGATACAGCACTGAGCAATGATGCCAGACCTTTCCT 480
Db 421 ACCTCGGAGGCTTCCTCGGAGTTTGGATACAGCACTGAGCAATGATGCCAGACCTTTCCT 480
Qy 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGATTTGGAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGATTTGGAAATTTG 540
Qy 541 GCAGATGTTTATAGCAATGCTTCCTGGCCCTGAATAGCACTGGCAGAGCAATTTGTATC 600
Db 541 GCAGATGTTTATAGCAATGCTTCCTGGCCCTGAATAGCACTGGCAGAGCAATTTGTATC 600
Qy 601 TCCTGTGAGTGCCCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTTACAGAAATCCGA 660
Db 601 TCCTGTGAGTGCCCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTTACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATGATGATTCCTGGAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATGATGATTCCTGGAAAGTATAAG 720
Qy 721 AGTATCTGGAATGGAATGTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
Db 721 AGTATCTGGAATGGAATGTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAAATCAGAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAAATCAGAA 840
Qy 841 GTAACCTCAGATGGCCCTCAGGCCATCATGGCTGCTCCTTTATTCATGTCTAATCAGCTC 900
Db 841 GTAACCTCAGATGGCCCTCAGGCCATCATGGCTGCTCCTTTATTCATGTCTAATCAGCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATTAAGGAGTAATGGCCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATTAAGGAGTAATGGCCATCAAT 960
Qy 961 CAGGACCCCTGGCAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTGGCAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAACCGGAGGAGATGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAACCGGAGGAGATGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTCGGTAAAGGAGTGGCCTGTATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTCGGTAAAGGAGTGGCCTGTATCCT
```

```
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTCGGTAAAGGAGTGGCCTGTATCCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAAAGGAGCTAGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAAAGGAGCTAGGCTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
Db 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
Qy 1261 ATGT 1264
Db 1261 ATGT 1264
```

RESULT 8

```
US-10-103-327-13
; Sequence 13, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; CURRENT APPLICATION NUMBER: US/10/103.327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-10-103-327-13
```

Query Match 99.8%; Score 1264; DB 5; Length 1284;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGGCTTCGCTTCCTGGCC 60

Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGGCTTCGCTTCCTGGCC 60

Qy 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGATGGCAAGGAGCGCT 120

Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGATGGCAAGGAGCGCT 120

Qy 121 ACCATGGGCTGGCTGCACTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180

Db 121 ACCATGGGCTGGCTGCACTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180

Qy 181 GATTCTCTGATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240

Db 181 GATTCTCTGATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240

Qy 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300

Db 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300

Qy 301 GATTGAGAGGAGCACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

Db 301 GATTGAGAGGAGCACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

Qy 361 GCTAATTATGTTACAGCAAGGAGCTGAGCTAGGATTTATGCAAGATTTGGAATATAA 420

Db 361 GCTAATTATGTTACAGCAAGGAGCTGAGCTAGGATTTATGCAAGATTTGGAATATAA 420

QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGCGACAGGG 780
DB 721 AGTAICTTGGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGCGACAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
DB 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCATTAATGACCTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCATTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGAGCTAAATTCGCCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGAGCTAAATTCGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGCTACCACTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGCTACCACTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAAACCGGAGAGATTTGGT 1080
DB 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAAACCGGAGAGATTTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCCCTGTAATCCT 1140
DB 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCCCTGTAATCCT 1140
QY 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGACATGTTTTCGCTCAGCTAGAAAACACA 1260
DB 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGACATGTTTTCGCTCAGCTAGAAAACACA 1260
QY 1261 ATGT 1264
DB 1261 ATGT 1264

RESULT 11
US-10-851-388-13
; Sequence 13, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-10-851-388-13
Query Match 99.8%; Score 1264; DB 8; Length 1284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGCTTGGCTTCGCTTCCTGGCC 60

DB 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGCTTGGCTTCGCTTCCTGGCC 60
QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGATGATGGCAAGGACGCT 120
DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGATGATGGCAAGGACGCT 120
QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCTGCAGGAAGAGCA 180
DB 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTTGACTGCTGCAGGAAGAGCA 180
QY 181 GATTCTTCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB 181 GATTCTTCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGGAAAGATGACAGGTTATGAGTACCTCTGCAATGATGACTGTTGATGGCTCCCCAAAGA 300
DB 241 TGGAAAGATGACAGGTTATGAGTACCTCTGCAATGATGACTGTTGATGGCTCCCCAAAGA 300
QY 301 GATTTCAGAAAGCAGACTTCAGGACAGCCCTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTTCAGAAAGCAGACTTCAGGACAGCCCTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATATGTTTACAGCAAAAGGACTGAAGCTAGGGAATTTATGCAAGATGTTGGAAATAAA 420
DB 361 GCTAATATGTTTACAGCAAAAGGACTGAAGCTAGGGAATTTATGCAAGATGTTGGAAATAAA 420
QY 421 ACCTGGCAGAGCTCCCTGGGAGTTTGGATACCTAGCACTGATGCTCCAGACACTTTGCT 480
DB 421 ACCTGGCAGAGCTCCCTGGGAGTTTGGATACCTAGCACTGATGCTCCAGACACTTTGCT 480
QY 481 GACTGGGAGTAGATCTCTAAATTTTCATGTTGTTACTGTGACAGTTTGGAAATTTG 540
DB 481 GACTGGGAGTAGATCTCTAAATTTTCATGTTGTTACTGTGACAGTTTGGAAATTTG 540
QY 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGTAATAGGACTGGCAGAAAGCATTTGTAC 600
DB 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGTAATAGGACTGGCAGAAAGCATTTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTTATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTTATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGAAAAGATATAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGAAAAGATATAAG 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
DB 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCATTAATGACCTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCATTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGAGCTAAATTCGCCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGAGCTAAATTCGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGCTACCACTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGCTACCACTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAAACCGGAGAGATTTGGT 1080
DB 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAAACCGGAGAGATTTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCCCTGTAATCCT 1140

Db 1081 GGACCTCGCTCTTATACCATGCGAGTGTCTCCCTGGGTAAAGGAGTGGCCGTGAATCCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGTAAAGAAAGCTAGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGTAAAGAAAGCTAGGCTTCTATGAATGGACT 1200
Qy 1201 TCAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAACACA 1260
Db 1201 TCAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAACACA 1260
Qy 1261 ATGT 1264
Db 1261 ATGT 1264

RESULT 12

US-10-984-389-13
; Sequence 13, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-10-984-389-13

Query Match 99.8%; Score 1264; DB 9; Length 1284;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTGGCTTCTCTGGCC 60
Db 1 ATGAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTGGCTTCTCTGGCC 60
Qy 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 121 ACCATGGGCTGGCTGCATCTGGAGGCTTTCATGTGCAACTTGTACTGGCAGGAAGGCA 180
Db 121 ACCATGGGCTGGCTGCATCTGGAGGCTTTCATGTGCAACTTGTACTGGCAGGAAGGCA 180
Qy 181 GATTCCTGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db 181 GATTCCTGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Qy 301 GATTCAGAGGAGCACTTCAGGCAAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAGAGGAGCACTTCAGGCAAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTTATGTTACAGCAAGAGCTGAAGCTAGGATTTATGCAAGATGTTGGAATAAA 420
Db 361 GCTAATTTATGTTACAGCAAGAGCTGAAGCTAGGATTTATGCAAGATGTTGGAATAAA 420

Qy 421 ACCTGGCAGGCTTCCCTGGGAGTTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGCTAAATTTTCATGCTTGTACTGTGACAGTTCGAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTTCATGCTTGTACTGTGACAGTTCGAAATTTG 540
Qy 541 GCAGATGTTTAAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAAAGCAATTTG 600
Db 541 GCAGATGTTTAAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAAAGCAATTTG 600
Qy 601 TCCTGTGAGTGGCTCTTTTATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTTATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGCGGAAATTTTCTGACATTTGATGATTCCTGAAAGATTAAG 720
Db 661 CAGTACTGCAATCACTGCGGAAATTTTCTGACATTTGATGATTCCTGAAAGATTAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAAATCAGCA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAAATCAGCA 840
Qy 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGGAGGTACAGCTTAGACAGGAGAGCACTTTGAAGTGGG 1020
Db 961 CAGGACCCCTTGGGCAAGGAGGTACAGCTTAGACAGGAGAGCACTTTGAAGTGGG 1020
Qy 1021 GAAAGCCTCTCTCAGGCTTACGCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGT 1080
Db 1021 GAAAGCCTCTCTCAGGCTTACGCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTTCCTTCCCTGGGTAAAGGAGTGGCCTGTAACTCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTCCTTCCCTGGGTAAAGGAGTGGCCTGTAACTCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAAAGCTAGGCTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAAAGCTAGGCTTCTATGAATGGACT 1200
Qy 1201 TCAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAACACA 1260
Db 1201 TCAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAACACA 1260
Qy 1261 ATGT 1264
Db 1261 ATGT 1264

RESULT 13

US-09-993-059-7
; Sequence 7, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059

```
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(1275)
US-09-993-059-7

Query Match      99.8%; Score 1263.4; DB 3; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTGGCTTCTCGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTGGCTTCTCGGCC 60

QY 61 CTCGTTTCTTGGGACATCTCTGGGCTAGAGCACTGGACATGGATTTGGCAAGGAGCCT 120
DB 61 CTCGTTTCTTGGGACATCTCTGGGCTAGAGCACTGGACATGGATTTGGCAAGGAGCCT 120

QY 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCA 180
DB 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCA 180

QY 181 GATTCCTGTCATCAGTGAAGGCTTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB 181 GATTCCTGTCATCAGTGAAGGCTTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 241 TGGAGGATGCAAGTTATGAGTACCTCTGCTGATTTGATGATGCTGGTCTCCCAAGA 300
DB 241 TGGAGGATGCAAGTTATGAGTACCTCTGCTGATTTGATGATGCTGGTCTCCCAAGA 300

QY 301 GATTCAGAGGAGGAGCTTCAGGAGAGCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCAGAGGAGGAGCTTCAGGAGAGCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

QY 361 GCTAATTTATGTTTACAGCAAAAGGAGCTGAAGCTAGGATTTATGCAAGATGTTTGGAAATAA 420
DB 361 GCTAATTTATGTTTACAGCAAAAGGAGCTGAAGCTAGGATTTATGCAAGATGTTTGGAAATAA 420

QY 421 ACTCGGAGGCTTCCTGGGAGTTTGGATGACTAGCAATGATGCCAGACCTTTGCT 480
DB 421 ACTCGGAGGCTTCCTGGGAGTTTGGATGACTAGCAATGATGCCAGACCTTTGCT 480

QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540

QY 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTAC 600
DB 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTAC 600

QY 601 TCTGTAGTGGCTCTTTATATGATGGCCCTTTCAAAGGCCAAATATACAGAAATCCGA 660
DB 601 TCTGTAGTGGCTCTTTATATGATGGCCCTTTCAAAGGCCAAATATATACAGAAATCCGA 660

QY 661 CAGTACTGCAATCAGCTGGGAAATTTGCTGACATTTGATGATTTCTGGGAAAGTATAAG 720
DB 661 CAGTACTGCAATCAGCTGGGAAATTTGCTGACATTTGATGATTTCTGGGAAAGTATAAG 720

QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTTGTGCGACAGGG 780
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTTGTGCGACAGGG 780

QY 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAAATCAGCA 840
DB 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAAATCAGCA 840

QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGGCTGCTCTTTTATTTCATGTCTAATGACCTC 900
```

```
DB 841 GFAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTTCATGCTTAATGACCTC 900

QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGAGCTAATTTGCCATCAAT 960
DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGAGCTAATTTGCCATCAAT 960

QY 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGGAGACAACTTTGAACTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGGAGACAACTTTGAACTGTGG 1020

QY 1021 GAAAGACCTCTCTCAGGCTTAGCCCTGAGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
DB 1021 GAAAGACCTCTCTCAGGCTTAGCCCTGAGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080

QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGAGTGGCCTGTAACTCT 1140
DB 1081 GGACCTCGCTCTTATACCATCGCATGCTTCCCTGGGTAAGAGAGTGGCCTGTAACTCT 1140

QY 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAGCTAGGGTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAGCTAGGGTTCTATGAATGGACT 1200

QY 1201 TCAGGTTTAAAGAGTCAATAAATCCCAAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
DB 1201 TCAGGTTTAAAGAGTCAATAAATCCCAAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260

QY 1261 ATGTA 1265
DB 1261 ATGCA 1265

RESULT 14
US-10-103-327-7
; Sequence 7, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(1275)
US-10-103-327-7

Query Match      99.8%; Score 1263.4; DB 5; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTGGCTTCTCGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTGGCTTCTCGGCC 60

QY 61 CTCGTTTCTTGGGACATCTCTGGGCTAGAGCACTGGACATGGATTTGGCAAGGAGCCT 120
DB 61 CTCGTTTCTTGGGACATCTCTGGGCTAGAGCACTGGACATGGATTTGGCAAGGAGCCT 120

QY 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCA 180
DB 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCA 180
```

QY 181 GATTCTCGATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db |||||
QY 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATGATGATCTGTTGATGGCTCCCAAGA 300
Db |||||
QY 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATGATGATCTGTTGATGGCTCCCAAGA 300
Db |||||
QY 301 GATTGAGAGGAGAGCTTCAGGAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db |||||
QY 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGGATTTATGAGATGTTGGAAATAAA 420
Db |||||
QY 421 ACCTGCGAGAGCTTCCTGGGAGTTTGGATCTAGCACTAGCACTGATGCCAGACCTTGGCT 480
Db |||||
QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTGTGTTACTGTGACAGTTCGAAAATTTG 540
Db |||||
QY 541 GCAGATGTTTAAAGCAGATGCTCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db |||||
QY 601 TCCTGTGATGGCTCTTTATATGATGGGCTTTTCAAAAGCCCAATTTACAGAAATCCGA 660
Db |||||
QY 661 CAGTACTGCAATCACTGCGAAATTTGCTGACATTTGATGATCTGCTGAAAGTATAAG 720
Db |||||
QY 721 AGTATCTTGGAGTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
Db |||||
QY 781 GGTGGAGTACCCAGATATGTTAGTGTGGAATCTTGGCTCAGCTGGAATCAGAA 840
Db |||||
QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Db |||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAGGAGCTAAATGCCATCAAT 960
Db |||||
QY 961 CAGGACCCCTGGGCAAGAGGATACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db |||||
QY 1021 GAAAGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGGAGATGGT 1080
Db |||||
QY 1081 GGACCTCGCTCTTATACCATCGCAGTGTGCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
Db |||||
QY 1141 GCCTGCTTATCAGACAGCTCTCCCTGTGAAAGAGAGCTAGGGTCTATGAAATGACT 1200
Db |||||
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGCAGCTGTTTGGCTTCAGCTAGAAACACA 1260
Db |||||

QY 1261 ATGTA 1265
Db |||||
QY 1261 ATGCA 1265

RESULT 15

US-10-602-219-7
; Sequence 7, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-7

Query Match 99.8%; Score 1263.4; DB 7; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
Db |||||
QY 61 CTCGTTCTCTGGGACATCCCTGGGCTAGACACTGGACAATGGATTGGCAAGGACGCT 120
Db |||||
QY 61 CTCGTTCTCTGGGACATCCCTGGGCTAGACACTGGACAATGGATTGGCAAGGACGCT 120
Db |||||
QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGAATGCTGCAGAGAGCA 180
Db |||||
QY 181 GATTCTGCACTAGTGAGAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db |||||
QY 181 GATTCTGCACTAGTGAGAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db |||||
QY 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATTCATGATCTGTTGATGGCTCCCAAGA 300
Db |||||
QY 241 TGGAGGATGAGGTTATGAGTACCTCTGCAATTCATGATCTGTTGATGGCTCCCAAGA 300
Db |||||
QY 301 GATTGAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCTCTCATGGATTCGCCAGCTA 360
Db |||||

```
Db      ||||| 301 GATT CAG AAGG CAG ACTT CAG CAG ACCCT CAG CGCTT TCCT CAT GGG ATTC GCC CAG CTA 360
QY      ||||| 361 GCT AAT TAT GTT CAC AAG CAA AGG ACT GAA GCT AGG GAT TTA TGC AGA GTT TGC GAA AATA AA 420
Db      ||||| 361 GCT AAT TAT GTT CAC AAG CAA AGG ACT GAA GCT AGG GAT TTA TGC AGA GTT TGC GAA AATA AA 420
QY      ||||| 421 ACC TGG CAG AGG CTT CCG TGG GAG TTT TGG ATA CTA GCA CATT GAT GCC CAG ACC TTT GCT 480
Db      ||||| 421 ACC TGG CAG AGG CTT CCG TGG GAG TTT TGG ATA CTA GCA CATT GAT GCC CAG ACC TTT GCT 480
QY      ||||| 481 GACT GGG AGT AGA TCT GCT AAAA TTT GAT GGT GTT GCT GCA CAG TTT TGG AAAA TTT G 540
Db      ||||| 481 GACT GGG AGT AGA TCT GCT AAAA TTT GAT GGT GTT GCT GCA CAG TTT TGG AAAA TTT G 540
QY      ||||| 541 GCAG ATG GTT ATA AAG CAC ATG TCT TGG CCG CCG TGA ATAG GACT GGG CAG AAG CAT TGT GTAC 600
Db      ||||| 541 GCAG ATG GTT ATA AAG CAC ATG TCT TGG CCG CCG TGA ATAG GACT GGG CAG AAG CAT TGT GTAC 600
QY      ||||| 601 TCC TGT GAG TGG CCG CTT TAT ATG TGG CCG CTT TCA AAA GCC CAA TTA TAC AGA AAT CCG A 660
Db      ||||| 601 TCC TGT GAG TGG CCG CTT TAT ATG TGG CCG CTT TCA AAA GCC CAA TTA TAC AGA AAT CCG A 660
QY      ||||| 661 CAG TACT GCA ATCA CT GGG CAA AAT TTT GCT GCA CATT GAT GAT TCCT TGG AAAA GTATAAAG 720
Db      ||||| 661 CAG TACT GCA ATCA CT GGG CAA AAT TTT GCT GCA CATT GAT GAT TCCT TGG AAAA GTATAAAG 720
QY      ||||| 721 AGT ATCT TGG ACT GGA CACT CTTTTT AA CAG GAG AAG AAT TTT GAT GTT GCT CGA CAG GG 780
Db      ||||| 721 AGT ATCT TGG ACT GGA CACT CTTTTT AA CAG GAG AAG AAT TTT GAT GTT GCT CGA CAG GG 780
QY      ||||| 781 GGT TGG AAT GAC CCA GAT ATG TTAG TAT GGG CAA CTT TGG CCG CAG CTG GAA T CAG CAA 840
Db      ||||| 781 GGT TGG AAT GAC CCA GAT ATG TTAG TAT GGG CAA CTT TGG CCG CAG CTG GAA T CAG CAA 840
QY      ||||| 841 GTA ACT CT CAG ATG GCG CCG CTT CAG GCG TAT CAT GCG TCT CTT TAT TCA TGT CTA ATG ACCTC 900
Db      ||||| 841 GTA ACT CT CAG ATG GCG CCG CTT CAG GCG TAT CAT GCG TCT CTT TAT TCA TGT CTA ATG ACCTC 900
QY      ||||| 901 CGA CAC ATC AGC CCG CTT CAA GCC AAA GCT CTT CAG GATA AGG ACG TAA TTGCC ATCAAT 960
Db      ||||| 901 CGA CAC ATC AGC CCG CTT CAA GCC AAA GCT CTT CAG GATA AGG ACG TAA TTGCC ATCAAT 960
QY      ||||| 961 CAG GAC CCG CTT GGG CAA GCA AGG GTAC CAG CTTAG ACG GAG ACAA CTTT GAA GTG TG 1020
Db      ||||| 961 CAG GAC CCG CTT GGG CAA GCA AGG GTAC CAG CTTAG ACG GAG ACAA CTTT GAA GTG TG 1020
QY      ||||| 1021 GAA CGAC CTTCT CAG GCTTAG CCG TGTAG CTTATG ATATAA CCG CAG GAG ATTGGT 1080
Db      ||||| 1021 GAA CGAC CTTCT CAG GCTTAG CCG TGTAG CTTATG ATATAA CCG CAG GAG ATTGGT 1080
QY      ||||| 1081 GGA CCG CCG CTTTAT ACCAT CCG AGT TGT TCC TGG GTAA AGG AGT GCG CTTGTAATCCT 1140
Db      ||||| 1081 GGA CCG CCG CTTTAT ACCAT CCG AGT TGT TCC TGG GTAA AGG AGT GCG CTTGTAATCCT 1140
QY      ||||| 1141 GCCT GCTT CAT CAC AGC CTTCC CCG TGTGA AAG GAG CTA GGG TTTCTATG ATGGACT 1200
Db      ||||| 1141 GCCT GCTT CAT CAC AGC CTTCC CCG TGTGA AAG GAG CTA GGG TTTCTATG ATGGACT 1200
QY      ||||| 1201 TCA AGGT TAA GAG TAC ATAAT CCA CAG GCA CTTG TTTT GCTT CAG CTAG AAAA CACA 1260
Db      ||||| 1201 TCA AGGT TAA GAG TAC ATAAT CCA CAG GCA CTTG TTTT GCTT CAG CTAG AAAA CACA 1260
QY      ||||| 1261 ATGTA 1265
Db      ||||| 1261 ATGCA 1265
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 08:30:05 ; Search time 284 Seconds
(without alignments)
2313.369 Million cell updates/sec

Title: US-10-602-219-11
Perfect score: 1266
Sequence: 1 atgcagctggaggaaccaga.....agctagaacaacaaatgtaa 1266

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4168288 seqs, 259477437 residues

Total number of hits satisfying chosen parameters: 8336576

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA_New.*

- 1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	37	2.9	2779	6	US-10-750-185-40681
C 2	36.6	2.9	1306	6	US-10-750-185-42255
C 3	35	2.8	321019	6	US-10-995-561-13204
C 4	34	2.7	1932	6	US-10-750-185-55363
C 5	33	2.6	1437	6	US-10-750-185-43877
C 6	32.8	2.6	1666	6	US-10-821-234-149
C 7	32.8	2.6	146656	7	US-11-121-086-68
C 8	32.8	2.6	188682	7	US-11-112-908-23
C 9	32.8	2.6	190882	7	US-11-121-086-69
C 10	32.4	2.6	1685	6	US-10-750-185-53897
C 11	32.4	2.5	86081	6	US-10-995-561-13246
C 12	32	2.5	1254	6	US-10-750-185-29504
C 13	31.8	2.5	201	6	US-10-995-561-29100
C 14	31.8	2.5	4847	6	US-10-750-185-60280
C 15	31.6	2.5	2787	6	US-10-841-129-3
C 16	31.6	2.5	1125000	6	US-10-995-561-13286
C 17	31.4	2.5	119160	7	US-11-121-086-12
C 18	31.4	2.5	161874	7	US-11-121-086-75
C 19	31.2	2.5	1522	6	US-10-750-185-37196
C 20	31	2.4	854	6	US-10-750-185-28311
C 21	31	2.4	1084	6	US-10-750-185-61883
C 22	31	2.4	1783	6	US-10-750-185-37071
C 23	31	2.4	168516	7	US-11-121-086-3

C 24	30.8	2.4	944	6	US-10-750-185-64685	Sequence 64685, A
C 25	30.8	2.4	1806	6	US-10-131-826A-121	Sequence 121, App
C 26	30.8	2.4	5234	6	US-10-821-234-357	Sequence 357, App
C 27	30.6	2.4	1131	6	US-10-750-185-54302	Sequence 54302, A
C 28	30.6	2.4	40394	6	US-10-995-561-13493	Sequence 13493, A
C 29	30.4	2.4	2103	6	US-10-750-185-42925	Sequence 42925, A
C 30	30.4	2.4	163162	7	US-11-121-086-66	Sequence 66, Appl
C 31	30.4	2.4	191091	7	US-11-121-086-60	Sequence 60, Appl
C 32	30.2	2.4	1805	6	US-10-750-185-39400	Sequence 39400, A
C 33	30.2	2.4	126552	7	US-11-121-086-1	Sequence 1, Appl
C 34	30.2	2.4	317876	6	US-10-995-561-13227	Sequence 13227, A
C 35	30	2.4	1267	6	US-10-750-185-43163	Sequence 43163, A
C 36	29.8	2.4	858	6	US-10-750-185-28504	Sequence 28504, A
C 37	29.8	2.4	2242	6	US-10-750-185-39271	Sequence 39271, A
C 38	29.8	2.4	3021	6	US-10-750-185-35062	Sequence 35062, A
C 39	29.8	2.4	3821	6	US-10-750-185-63028	Sequence 63028, A
C 40	29.8	2.4	7980	6	US-10-509-921-4	Sequence 4, Appl
C 41	29.8	2.4	7980	6	US-10-509-921-5	Sequence 5, Appl
C 42	29.8	2.4	7989	6	US-10-509-921-2	Sequence 2, Appl
C 43	29.8	2.4	7989	6	US-10-509-921-6	Sequence 6, Appl
C 44	29.8	2.4	7989	6	US-10-509-921-13	Sequence 13, Appl
C 45	29.8	2.4	7989	6	US-10-509-921-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-10-750-185-40681/c
; Sequence 40681, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10750185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40681
; LENGTH: 2779
; TYPE: DNA
; ORGANISM: Bovine 19866880781246
US-10-750-185-40681

Query Match 2.9%; Score 37; DB 6; Length 2779;
Best Local Similarity 52.2%; Pred. No. 0.082;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
Qy 631 TTTCAAAGCCCAATTATACAGAAATCCGACAGTACTGCAATCACTGGCGAAATTTTGT 690
Db 2693 TTTTAGACTCCCTGTTTGATTAGGTTTGTCTTATCTTATGCGAGTTCTACTGTTT 2634
Qy 691 GACATTGATGATCTCTGGAAAAGTATAAAGAGTATCTTGGACTGGACATCTTTTAAACGAG 750
Db 2633 GACTCTTGGGCTCAGGATACACTGTAAACTTCAGGAGGCTTGTGTATCTGTATACCAG 2574
Qy 751 GAGAGATTGTTGATGTTGCTGGACCGGGGGTTGGA 787
Db 2573 CAAAGGAGTATTGATGTTGCTGTACCAAGTTTGTAA 2537

RESULT 2
US-10-750-185-42255
; Sequence 42255, Application US/10750185

; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 42255
; LENGTH: 1306
; TYPE: DNA
; ORGANISM: Bovine 19866880955804
US-10-750-185-42255

Query Match 2.9%; Score 36.6; DB 6; Length 1306;
Best Local Similarity 53.1%; Pred. No. 0.067;
Matches 78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 715 ATAAGAGATATCTTGACCTGGAGCATCTTTTAACGAGGAGAGAAATTTGTGATGTTGCTGGA 774
DB 584 ATGCAGGTAAATAATATAGACAATTTAGTCAAGAGAGCAATTTGTTGTGAAGAAATTC 643
QY 775 CCAGGGGTTGAATGACCCAGCATATGTTAGTGATTGGCAACTTTGGCCTCAGCTGGAAT 834
DB 644 CCAACCCCTGGAACCTTGATGATTTCTGTAATATAGGCGAGCTGGCCCTTCCTTGCCTT 703
QY 835 CAGCAAGTAACCTCAGATGCCCTCTGG 861
DB 704 CTGTCCCTGCCTCAGGTGGTCCCTGG 730

RESULT 3
US-10-995-561-13204/c
; Sequence 13204, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13204
; LENGTH: 321019
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(321019)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13204

Query Match 2.8%; Score 35; DB 6; Length 321019;
Best Local Similarity 49.1%; Pred. No. 11;
Matches 86; Conservative 2; Mismatches 87; Indels 0; Gaps 0;
QY 915 TCAAGCAAGGTCCTCTTCAGATTAAGGACGTAATTTGCCATCATCAGGACCCCTTGGG 974
DB 301377 TCACACAGAGGCTGAGTACAGAACACRAAGATTTGTAAAGACAGGATGCTTAAG 301318
QY 975 CAAGCAAGGGTACCAGCTTAGACAGGAGACAACCTTTGAAGTGTGGGAACGACCTCTCTC 1034

DB 301317 CATGAGGCTGGCTGCCTTTGTGAGGCAGTGAACATGAGTCCCTGGGAAGAAACCTCTT 301258
QY 1035 AGGCTTAGCCTGGGCTGTAGCTATCATATAAACCGGACGAGAGATTGTTGGACCTCGC 1089
DB 301257 AGCAGAAGCTCAACACGACGCTGTCAAAGATCTGCAGGTGCACCTGGGAGCTGGGC 301203
RESULT 4
US-10-750-185-55363/c
; Sequence 55363, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 55363
; LENGTH: 1932
; TYPE: DNA
; ORGANISM: Bovine 19866880928582
US-10-750-185-55363

Query Match 2.7%; Score 34; DB 6; Length 1932;
Best Local Similarity 57.5%; Pred. No. 0.7;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 655 ATCCGACAGTACTGCAATCACTGGCGAAATTTTGTGACATTGATGATTCCTGGAAAAGT 714
DB 1081 ATCTGAGATTTCTGAGATGATTTTCCAAAATTTATTTGGAGAAAGGATTCCTGAAGAATT 1022
QY 715 ATAAAGAGTATCTTGGACTGGACATCTTTTAAACGAGGAGAGAATTG 760
DB 1021 AGTGAATTTCTTGATGTACATACTTTTATCGACATTATAATTG 976

RESULT 5
US-10-750-185-43877/c
; Sequence 43877, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 43877
; LENGTH: 1437
; TYPE: DNA
; ORGANISM: Bovine 19866880874544
US-10-750-185-43877


```
Query Match      2.6%; Score 33; DB 6; Length 1437;
Best Local Similarity 53.5%; Pred. No. 1.3;
Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 835 CAGCAAGTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAAT 894
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 458 CAGGAATTTATTTGGTGTGGCTGACTGGGATATCCAGGAGGAGCTGGCTCAGGAATTC 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 895 GACCTCCGACACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAGGATAAGGACGTAAATGCC 954
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 398 GAGCCAGAGCCCTTCTGGCCCTCCCTCTCATCTCTCCCGTGTACTAGATGCTCTTTGGA 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 955 ATCAATCAG 963
    ||| ||| |||
Db 338 ATGGAACAG 330
    ||| ||| |||

RESULT 6
US-10-821-234-149
; Sequence 149, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 149
; LENGTH: 1666
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-149

Query Match      2.6%; Score 32.8; DB 6; Length 1666;
Best Local Similarity 64.5%; Pred. No. 1.6;
Matches 49; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 156 CAACCTTGACTCCAGGAGAGCCAGATTCTCTGCATCATGAGAGCTCTTCATGGAGAT 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 CAGTCTTGCCCTCCCTGGATATCTCCGATTACTCCAAATGTAAGAGATCATGATTGAGAG 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 216 GGCAGAGCTCATGGTC 231
    ||| ||| ||| |||
Db 298 AGGAGAGCTTTTCTC 313
    ||| ||| ||| ||| |||

RESULT 7
US-11-121-086-68
; Sequence 68, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-0000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 68
; LENGTH: 146656
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-68

Query Match      2.6%; Score 32.8; DB 7; Length 188682;
Best Local Similarity 59.8%; Pred. No. 43;
Matches 55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 256 TATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGAGATTTCAGAGGCGAGA 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139854 TATGCTATCTCTGGATTATCACTGTAGTTGGGGAATAAGGAATCTGACTGCCAGA 139795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 316 CTTGAGCAGACCTCCAGCGCTTTCTCTCATCG 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139794 TCTGAGACATAGTCCACCCCTTTTGCACTGGG 139763
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-11-121-086-69
; Sequence 69, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-0000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 69
; LENGTH: 190882
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-69

Query Match      2.6%; Score 32.8; DB 7; Length 146656;
Best Local Similarity 58.0%; Pred. No. 37;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 644 ATTATACAGAATCCGACAGTACTGCAATCACTCGCGAATTTTCTGACATTGATGATT 703
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134039 ATTACAGCATATGCCAAAGTACTAGGCTCACTTGTCTAATTTTAGCACCCCAATGATGATT 134098
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 704 CCGTGAAGATATAAGAGATATCTTGGACTGGACATCTTT 743
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134099 CTTCTTATATTAATCACTTATCTCTTAATAGAAATTTATT 134138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
US-11-112-908-23/c
; Sequence 23, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 23
; LENGTH: 188682
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-23
```

US-11-121-086-69

Query Match 2.6%; Score 32.8; DB 7; Length 190882;
Best Local Similarity 58.0%; Pred. No. 44; Mismatches 42; Indels 0; Gaps 0;
Matches 58; Conservative 0;

QY 644 ATTATACAGAAATCCGACAGTACTGCAATCACTGGCGAAATTTTGTGACATTGATGATT 703
|||||
Db 22276 ATTACAGGATATGCCAAGTACTAGGCTCACTTGTCTAATTTTAGCACCAATGATGATT 22335
|||||

QY 704 CTGGAAAAGTAAAGAGTATCTTGGACTGGACATCTTTT 743
|||||
Db 22336 CTTCTTATATTAATCACTTATCTTAAATAGAAATTAAT 22375
|||||

RESULT 10

US-10-750-185-53897
; Sequence 53897, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 53897
; LENGTH: 1685
; TYPE: DNA
; ORGANISM: Bovine 19866880964900
US-10-750-185-53897

Query Match 2.6%; Score 32.4; DB 6; Length 1685;
Best Local Similarity 60.0%; Pred. No. 2.3; Mismatches 36; Indels 0; Gaps 0;
Matches 54; Conservative 0;

QY 127 GCGTGCCTGCACTGGAGGCGCTTCATGTCGAACCTTGACTGCCAGGAGGCGAGATTCC 186
|||||
Db 849 GCGTTCTGCTGATGGCTCAGTTTATGAGCCACGTTTAAAGGAGAGGAGAAACCAACATTTT 908
|||||

QY 187 TGCATCAGTGAGAAGCTCTTCATGGAGATG 216
|||||
Db 909 TGTTCCTTGAGTAACTCGTTATAGATTGG 938
|||||

RESULT 11

US-10-995-561-13246/c
; Sequence 13246, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13246
; LENGTH: 86081
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13246

Query Match 2.5%; Score 32.2; DB 6; Length 86081;
Best Local Similarity 48.0%; Pred. No. 41; Mismatches 82; Conservative 3; Mismatches 86; Indels 0; Gaps 0;

QY 104 GATTGGCAAGGACGCTTACCATGGCTGGCTGCACCTGGAGGCGCTTCATGTGCAACCTTG 163
|||||
Db 68847 GGTGGGCAAGGCTTTTGACCTTGGCCCTGCGGTAGTGGCGTGAGGTACCTG 68788
|||||

QY 164 ACTGCCAGGAAGACGAGATTCTTCATCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGC 223
|||||
Db 68787 CGTCTGGGGTYGGACTGAGGGGACATCAGCATGTTGTTCTCCAGGGCGGCTGCARGTG 68728
|||||

QY 224 TCATGCTCTCAGAAGGCTGGAAGGATGCAGGTTATGAGTACCTCTGCAATTG 274
|||||
Db 68727 GCAGGCTGCCATCCAGCTGYAAATGCACCTAGCAGGGGCCCTGGCCCTG 68677
|||||

RESULT 12

US-10-750-185-29504
; Sequence 29504, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 29504
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Bovine 19866880446821
US-10-750-185-29504

Query Match 2.5%; Score 32; DB 6; Length 1254;
Best Local Similarity 46.4%; Pred. No. 2.5; Mismatches 120; Indels 0; Gaps 0;
Matches 104; Conservative 0;

QY 215 TGGCAGAGCTCATGGTCTCAGAAGGCTGGAAGGATGCGAGGTTATGAGTACCTCTGCAATTG 274
|||||
Db 440 TGGAGATCCCATGGACAGAGGAGCGCTGGCAGGCTACAGTCCATGGTGGTAAAGAAATTG 499
|||||

QY 275 ATGACTGTTGGATGGCTCCCAAGAGATTTCAGAAGGAGACTTCAGGAGACCCCTCAGC 334
|||||
Db 500 GACACAATTTAGTGAATAAACAATGATGCAGAAAAGCCATAATACATTTTGGAGAGAACT 559
|||||

QY 335 GCTTTCTCTATGGGATTCGCCAGCTAGCTAATTATGTTTCACAGAAAGGAGACTGAAGCTAG 394
|||||
Db 560 GGTGAGAGAGGAGGAGCAATAAATGTTTAAATAATATTCATGGCAACATCTCCGAGCAAA 619
|||||

QY 395 GGATTTATGCGAGATGTTGGAATAAAACCTGCGCAGGCTTCCT 438
|||||
Db 620 TTTAAATAATAAATAAAGTAAATAGACTCAATTTCTGCTACTCT 663
|||||

RESULT 13

US-10-995-561-29100/c
; Sequence 29100, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 24, 2005, 08:33:27 ; Search time 187 Seconds
(without alignments)
989.190 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MQLNPEHLGICALALRFLA.....RLRSHINPTGVLLQLENTM 421

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2314	100.0	421	5	Aae28210 Human rGA
2	2314	100.0	421	7	ADD84751 Human alp
3	2314	100.0	421	8	ADJ88277 Human WT
4	2314	100.0	421	8	ADM48681 Human wil
5	2314	100.0	421	8	ADU66916 Human alp
6	2314	100.0	421	9	Aea27445 Human alp
7	2314	100.0	425	9	AEA27472 Human alp
8	2314	100.0	427	5	Aae28211 Human rGA
9	2314	100.0	427	7	ADD84753 Human alp
10	2314	100.0	427	8	ADJ88279 Human WT
11	2314	100.0	427	8	ADM48683 Human wil
12	2314	100.0	427	8	ADU66918 Human alp
13	2314	100.0	427	9	AEA27447 Human alp
14	2314	100.0	429	2	AAR53765 Alpha-gal
15	2314	100.0	429	2	AAR63234 Human alp
16	2314	100.0	429	2	AAR70207 Alpha-gal
17	2314	100.0	429	5	Aae28206 Human WT
18	2314	100.0	429	7	ADf76604 Novel hum
19	2314	100.0	429	8	ADM49740 Human alp
20	2314	100.0	429	8	ADP24703 PRO poly
21	2314	100.0	429	8	ADU18066 Human can
22	2314	100.0	429	8	ADU74416 Human alp
23	2314	100.0	429	9	ADY19459 PRO poly
24	2314	100.0	429	9	ADY17095 PRO poly

25	2314	100.0	429	9	AEA27471 Human alp
26	2314	100.0	429	9	AEb43188 Human alp
27	2314	100.0	430	7	ABM79001 Human alp
28	2314	100.0	431	5	Aae28209 Human rGA
29	2314	100.0	431	7	ADD84749 Human alp
30	2314	100.0	431	8	ADJ88275 Human WT
31	2314	100.0	431	8	ADM48679 Human wil
32	2314	100.0	431	8	ADU66914 Human alp
33	2314	100.0	431	9	AEA27443 Human alp
34	2314	100.0	435	5	Aae28207 Human WT
35	2314	100.0	435	7	ADD84745 Human alp
36	2314	100.0	435	8	ADJ88271 Human WT
37	2314	100.0	435	8	ADM48675 Human wil
38	2314	100.0	435	8	ADU66910 Human alp
39	2314	100.0	435	9	AEA27439 Human alp
40	2314	100.0	478	2	RAY48575 Human bre
41	2311	99.9	429	2	AAR07305 Precursor
42	2309	99.8	424	5	Aae28208 Human rGA
43	2309	99.8	424	7	ADD84747 Human alp
44	2309	99.8	424	8	ADJ88273 Human WT
45	2309	99.8	424	8	ADM48677 Human wil

ALIGNMENTS

RESULT 1
AAE28210
ID AAE28210 standard; protein; 421 AA.
XX
AC AAE28210;
XX
DT 27-DEC-2002 (first entry)
XX
DE Human rGAL-8 protein.
XX
KW Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-8.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Cleavage-site 401..402
FT /note= "CTPP cleavage site"
XX
FN US2002088024-A1.
XX
PD 04-JUL-2002.
XX
PF 13-NOV-2001; 2001US-00993059.
XX
PR 26-JUL-2000; 2000US-00626127.
XX
PA (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX
PI Garger SJ, Turpen TH, Kumagai MH;
XX
DR WPI; 2002-681656/73.
DR N-PSDB; AAD45223.
XX
PT Novel human alpha-galactosidase polypeptide useful for treating lysosomal storage diseases.
XX
PS Claim 7; Page 44; 88pp; English.
XX
CC The invention relates to human alpha-galactosidase truncated at the carboxy terminus and the production of enzymatically active recombinant human and animal lysosomal enzymes. The invention is useful for producing CC lysosomal enzymes for treating lysosomal storage diseases, producing CC altered or mutated proteins, enzymatically active or otherwise, to serve as precursors or substrates for further in vivo or in vitro processing to

CC a specialised industrial form for research or therapeutic uses, to
CC produce more effective therapeutic enzyme, for producing antibodies
CC against lysosomal enzymes for medical diagnostic use, and in any
CC commercial process that involves substrate hydrolysis. The present
CC sequence is human rGAL-8 protein
XX
SQ Sequence 421 AA;

Query Match 100.0%; Score 2314; DB 5; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNCDCQEP 60
DB 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNCDCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQORDSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQORDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTADFADGVDLKFDGCGYCDSENL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTADFADGVDLKFDGCGYCDSENL 180
QY 181 ADGYKHSALNRTGRSIVYCEWPLYMPPQKPNYTEIROYCNHWRNFADIDDSWKSITK 240
DB 181 ADGYKHSALNRTGRSIVYCEWPLYMPPQKPNYTEIROYCNHWRNFADIDDSWKSITK 240
QY 241 SILDWTSPNQRIVDVAGPGWNPDMVLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSPNQRIVDVAGPGWNPDMVLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWRPLSGLAWAVAMINRQETG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWRPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421

RESULT 2
ADD84751
ID ADD84751 standard; protein; 421 AA.
XX
AC ADD84751;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human alpha-galactosidase rGAL-8 polypeptide.
XX
KW Human; alpha-galactosidase; rGAL-4; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease; enzyme.
XX
OS Homo sapiens.
XX
PN US2003106095-A1.
XX
PD 05-JUN-2003.
XX
PF 20-MAR-2002; 2002US-00103327.
XX
PR 26-JUL-2000; 2000US-00626127.
XX
PR 13-NOV-2001; 2001US-00993059.
XX
XX (GARG/) GARGER S J.
PA (TUREP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX

PI Garger SJ, Turpen TH, Kumagai MH;
XX
DR WPI; 2003-801257/75.
DR N-PSDB; ADD84750.
XX
PT New polynucleotide for producing active recombinant human and animal
PT lysosomal enzymes in a plant expression system that can be used in enzyme
PT replacement therapy.
XX
PS Claim 7; SEQ ID NO 12; 77pp; English.
XX
CC The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for
CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents a human alpha-galactosidase derivative
CC polypeptide of the invention.
XX
SQ Sequence 421 AA;

Query Match 100.0%; Score 2314; DB 7; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNCDCQEP 60
DB 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNCDCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQORDSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQORDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTADFADGVDLKFDGCGYCDSENL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPFGSGFYDDAQTADFADGVDLKFDGCGYCDSENL 180
QY 181 ADGYKHSALNRTGRSIVYCEWPLYMPPQKPNYTEIROYCNHWRNFADIDDSWKSITK 240
DB 181 ADGYKHSALNRTGRSIVYCEWPLYMPPQKPNYTEIROYCNHWRNFADIDDSWKSITK 240
QY 241 SILDWTSPNQRIVDVAGPGWNPDMVLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSPNQRIVDVAGPGWNPDMVLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWRPLSGLAWAVAMINRQETG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWRPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421

RESULT 3
ADJ88277
ID ADJ88277 standard; protein; 421 AA.
XX
AC ADJ88277;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human WT rGAL-8 (galactosidase).
XX
KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease; human; enzyme.
XX
OS Homo sapiens.
XX

```

PN US2004016021-A1.
XX
PD 22-JAN-2004.
XX
PF 23-JUN-2003; 2003US-00602219.
XX
PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
FI Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-108227/11.
XX
DR N-PSDB; ADJ88276.
XX
PT New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
PS Claim 7; SEQ ID NO 12; 71pp; English.
XX
CC The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
CC protein.
XX
XX Sequence 421 AA;
XX
Query Match 100.0%; Score 2314; DB 8; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCNLCDCQEEP 60
DB 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCNLCDCQEEP 60
QY 61 DSCISEKLFMEAEIWMVSEGKADAGEYVICDDCWAPQDSEGRLOADPQRFPHGIQRL 120
DB 61 DSCISEKLFMEAEIWMVSEGKADAGEYVICDDCWAPQDSEGRLOADPQRFPHGIQRL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGFPFGSGYDYDIDAQTFADWGVLLKPDGVCYCDLSNLL 180
DB 121 ANYVHSGKGLGIYADVGNKTCAGFPFGSGYDYDIDAQTFADWGVLLKPDGVCYCDLSNLL 180
QY 181 ADGYKXMSIALNRTGRSIVYSCWELPMYPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGYKXMSIALNRTGRSIVYSCWELPMYPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTSFNOERIVDVAGPGWNPDMVLVGNFGLSWNOQVQTMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNOERIVDVAGPGWNPDMVLVGNFGLSWNOQVQTMALWAIMAAPLFMSNDL 300
QY 301 RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPFLSGLAWAVAMINRQEI 360

```

```
CC invention.
XX
SQ Sequence 421 AA;
Query Match 100.0%; Score 2314; DB 8; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPHELGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLRNPHELGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPODSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPODSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYYDIDAQTFADWGVLLKFDGICYDLSLENL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYYDIDAQTFADWGVLLKFDGICYDLSLENL 180
QY 181 ADGYKHSALNLRTRGRSIVYSCWPLVYMWPPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGYKHSALNLRTRGRSIVYSCWPLVYMWPPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTSTFNOERIVDVAGPGWNPDMVLVGNFGLSWNQOVTOMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSTFNOERIVDVAGPGWNPDMVLVGNFGLSWNQOVTOMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
DB 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
QY 361 GPRSTYTIASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
DB 361 GPRSTYTIASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421
RESULT 5
ADU66916
ID ADU66916 standard; protein; 421 AA.
XX
AC ADU66916;
XX
DT 10-FEB-2005 (first entry)
XX
DE Human alpha-galactosidase protein #5.
XX
KW Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human.
XX
OS Homo sapiens.
XX
PN US2004234516-A1.
XX
PD 25-NOV-2004.
XX
PF 21-MAY-2004; 2004US-00851388.
XX
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
PR 20-MAR-2002; 2002US-00103327.
XX
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2004-821274/81.
XX
DR N-PSDB; ADU66915.
XX
PT A pharmaceutical composition comprising a lysosomal enzyme, useful for
PT enzyme replacement therapy for the treatment of lysosomal storage
PT diseases, such as Fabry's disease.
XX
PS Disclosure; SEQ ID NO 12; 88pp; English.
XX
CC The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase enzymes having a post-translational modification provided
CC by the plant expression system. The invention is useful in enzyme
CC replacement therapy for treating lysosomal storage diseases such as
CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase protein.
XX
SQ Sequence 421 AA;
Query Match 100.0%; Score 2314; DB 8; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPHELGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLRNPHELGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPODSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPODSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYYDIDAQTFADWGVLLKFDGICYDLSLENL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYYDIDAQTFADWGVLLKFDGICYDLSLENL 180
QY 181 ADGYKHSALNLRTRGRSIVYSCWPLVYMWPPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGYKHSALNLRTRGRSIVYSCWPLVYMWPPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTSTFNOERIVDVAGPGWNPDMVLVGNFGLSWNQOVTOMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSTFNOERIVDVAGPGWNPDMVLVGNFGLSWNQOVTOMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
DB 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
QY 361 GPRSTYTIASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
DB 361 GPRSTYTIASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQLENT 420
QY 421 M 421
DB 421 M 421
RESULT 6
AEA27445
ID AEA27445 standard; protein; 421 AA.
XX
AC AEA27445;
XX
DT 11-AUG-2005 (first entry)
XX
DE Human alpha-galactosidase protein, rGAL-8, SEQ ID NO: 12.
XX
XX Gauchers disease; metabolic; neurological disease; niemann pick disease;
XX genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
XX antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
```


KW enzyme.
XX Homo sapiens.
XX US2005125859-A1.
XX 09-JUN-2005.
XX 08-NOV-2004; 2004US-00984389.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX 20-MAR-2002; 2002US-00103327.
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2005-404004/41.
XX N-PSDB; AEA27444.
XX New isolated polypeptides useful for producing lysosomal enzymes in
XX plants to be utilized in enzyme replacement therapy or for the
XX therapeutic treatment of human or animal lysosomal storage diseases, e.g.
XX Gaucher's disease.
XX Claim 28; SEQ ID NO 12; 88pp; English.
XX The present invention relates to the production of human and animal
XX lysosomal enzymes in plants by a transient plant expression system. The
XX invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX galactosidase (Gal) enzymes having a post-translational modification
XX provided by the plant expression system. The invention is useful in
XX enzyme replacement therapy for treating lysosomal storage diseases such
XX as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
XX disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX also useful in researches for developing new approaches to medical
XX treatment of lysosomal storage diseases and in industrial processes
XX involving enzymatic substrate hydrolysis. The present sequence is the
XX human alpha-galactosidase protein.
XX Sequence 421 AA;
Query Match 100.0%; Score 2314; DB 9; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPHELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLRNPHELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEILMVSEGWKADGVEYLCIDDCWMAQRDSEGLQADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEILMVSEGWKADGVEYLCIDDCWMAQRDSEGLQADPQRFPHGIRQL 120
QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYGYDIDAQTFADGWDLKFKDCYDLSLENL 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGFYGYDIDAQTFADGWDLKFKDCYDLSLENL 180
QY 181 ADGQKMSLALNRTGRSIVYSCWPLVYWPQKPNYTEIRQYCNHWRNFADIDSWKSIK 240
DB 181 ADGQKMSLALNRTGRSIVYSCWPLVYWPQKPNYTEIRQYCNHWRNFADIDSWKSIK 240
QY 241 SILDWTSTNQRERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSTNQRERIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIADINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQIG 360
DB 301 RHISPOAKALLQDKDVIADINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQIG 360
QY 361 GPRSYTTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
DB 361 GPRSYTTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420

Db 361 GPRSYTTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
QY 421 M 421
Db 421 M 421
RESULT 7
AEA27472
ID AEA27472 standard; protein; 425 AA.
XX AEA27472;
XX 11-AUG-2005 (first entry)
XX Human alpha-galactosidase protein, rGAL-4, SEQ ID NO: 8 #2.
XX Gauchers disease; metabolic; neurological disease; niemann pick disease;
XX genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
XX antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
XX enzyme.
XX Homo sapiens.
XX US2005125859-A1.
XX 09-JUN-2005.
XX 08-NOV-2004; 2004US-00984389.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX 20-MAR-2002; 2002US-00103327.
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2005-404004/41.
XX N-PSDB; AEA27440.
XX New isolated polypeptides useful for producing lysosomal enzymes in
XX plants to be utilized in enzyme replacement therapy or for the
XX therapeutic treatment of human or animal lysosomal storage diseases, e.g.
XX Gaucher's disease.
XX Disclosure; SEQ ID NO 8; 88pp; English.
XX The present invention relates to the production of human and animal
XX lysosomal enzymes in plants by a transient plant expression system. The
XX invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX galactosidase (Gal) enzymes having a post-translational modification
XX provided by the plant expression system. The invention is useful in
XX enzyme replacement therapy for treating lysosomal storage diseases such
XX as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
XX disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX also useful in researches for developing new approaches to medical
XX treatment of lysosomal storage diseases and in industrial processes
XX involving enzymatic substrate hydrolysis. The present sequence is the
XX human alpha-galactosidase protein. Note: The present sequence is the SEQ
XX ID NO: 8 which is shown in page 36-38 of the specification. This sequence
XX differs from the SEQ ID NO: 8 given in the sequence listing (see
XX AEA27441).
XX Sequence 425 AA;
Query Match 100.0%; Score 2314; DB 9; Length 425;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPHELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLRNPHELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCNLDCCQEP 60

PA (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2003-801257/75.
DR N-PSDB; ADD84752.
DR
XX New polynucleotide for producing active recombinant human and animal
XX lysosomal enzymes in a plant expression system that can be used in enzyme
XX replacement therapy.
XX
XX Claim 7; SEQ ID NO 14; 77pp; English.
XX
XX The invention relates to human alpha-galactosidase derivatives and the
XX nucleic acids encoding them. The polypeptides are used in a method for
XX producing active recombinant human and animal lysosomal enzymes in a
XX plant expression system. The enzymes can be used in enzyme replacement
XX therapy for the therapeutic treatment of human and animal lysosomal
XX diseases. This sequence represents a human alpha-galactosidase derivative
XX polypeptide of the invention.
XX
XX Sequence 427 AA;
SQ
Query Match 100.0%; Score 2314; DB 7; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWDAGYEYLCIDDCWMAQORDSEGRLOADPQPFPHGIROL 120
DB 61 DSCISEKLFMEAEMLVSEGWDAGYEYLCIDDCWMAQORDSEGRLOADPQPFPHGIROL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPPGSGFYGYDIDAQTFADGWVLLKFDGCGYCDLSLENL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPPGSGFYGYDIDAQTFADGWVLLKFDGCGYCDLSLENL 180
QY 181 ADGYKHSALNRTGRSIVYSCWPLYNMFPQKPNYTEIRQYCNHWRNFADIDDSWSIK 240
DB 181 ADGYKHSALNRTGRSIVYSCWPLYNMFPQKPNYTEIRQYCNHWRNFADIDDSWSIK 240
QY 241 SILDWTSENOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSENOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYLQKQGNFVWERPLSGLAWAMINRQIG 360
DB 301 RHISPOAKALLQDKDVIAINODPLGKQGYLQKQGNFVWERPLSGLAWAMINRQIG 360
QY 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKKLGFYEWTSRLSHINPTGTVLQLENT 420
DB 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKKLGFYEWTSRLSHINPTGTVLQLENT 420
QY 421 M 421
DB 421 M 421
RESULT 10
ID ADJ88279
XX ADJ88279 standard; protein; 427 AA.
XX
XX AC ADJ88279;
XX
XX DT 06-MAY-2004 (first entry)
XX
XX DE Human WT rGAL-8R (galactosidase).
XX
XX KW Galactosidase; GAL; gene therapy; lysosomal storage disease;

KW Fabry's disease; Gaucher's disease; human; enzyme.
XX
XX OS Homo sapiens.
XX
XX US2004016021-A1.
XX
XX PD 22-JAN-2004.
XX
XX PF 23-JUN-2003; 2003US-00602219.
XX
XX PR 26-FEB-1988; 88US-00160766.
XX PR 26-FEB-1988; 88US-00160771.
XX PR 15-JUL-1988; 88US-00219279.
XX PR 17-FEB-1989; 89US-00310881.
XX PR 05-MAY-1989; 89US-00347637.
XX PR 08-JUN-1989; 89US-00363138.
XX PR 22-OCT-1990; 90US-00600244.
XX PR 16-JAN-1991; 91US-00841617.
XX PR 26-JUL-1991; 91US-00737899.
XX PR 01-AUG-1991; 91US-00739143.
XX PR 31-JUL-1992; 92US-00923692.
XX PR 30-DEC-1992; 92US-00997733.
XX PR 29-DEC-1993; 93US-00176414.
XX PR 19-JAN-1994; 94US-00184237.
XX PR 14-OCT-1994; 94US-00324003.
XX PR 21-MAY-1999; 99US-00316572.
XX PR 26-JUL-2000; 2000US-00626127.
XX PR 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
XX (POGU/) POGUE G P.
XX (ERWI/) ERWIN R L.
XX (GRIL/) GRILL L K.
XX
XX Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-108227/11.
XX N-PSDB; ADJ88278.
XX
XX New lysosomal enzymes, useful in treating human and animal lysosomal
XX storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
XX Claim 7; SEQ ID NO 14; 71pp; English.
XX
XX The invention relates to nucleotide encoding galactosidase (GAL). The
XX invention is useful in gene therapy. The polynucleotides and polypeptides
XX are useful in treating human and animal lysosomal storage diseases, e.g.
XX Fabry's disease and Gaucher's diseases. The present sequence is human GAL
XX protein.
XX
XX Sequence 427 AA;
SQ
Query Match 100.0%; Score 2314; DB 8; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
DB 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWDAGYEYLCIDDCWMAQORDSEGRLOADPQPFPHGIROL 120
DB 61 DSCISEKLFMEAEMLVSEGWDAGYEYLCIDDCWMAQORDSEGRLOADPQPFPHGIROL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPPGSGFYGYDIDAQTFADGWVLLKFDGCGYCDLSLENL 180
DB 121 ANYVHSKGLGIYADVGNKTCAGPPGSGFYGYDIDAQTFADGWVLLKFDGCGYCDLSLENL 180
QY 181 ADGYKHSALNRTGRSIVYSCWPLYNMFPQKPNYTEIRQYCNHWRNFADIDDSWSIK 240
DB 181 ADGYKHSALNRTGRSIVYSCWPLYNMFPQKPNYTEIRQYCNHWRNFADIDDSWSIK 240
QY 241 SILDWTSENOERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300

```
|||||
Db 241 SILDWTSTFQRIIVDVAGPGWNDPMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
QY 361 GPRSYTIAVASIGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
Db 361 GPRSYTIAVASIGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 11
ADM48683
ID ADM48683 standard; protein; 427 AA.
XX
AC ADM48683;
XX
DT 03-JUN-2004 (first entry)
XX
DE Human wild type rGAL-8R protein.
XX
KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW enzyme.
XX
OS Homo sapiens.
XX
PN US2004023281-A1.
XX
PD 05-FEB-2004.
XX
PF 23-JUN-2003; 2003US-00602220.
XX
PR 26-FEB-1988; 88US-00160766.
XX
PR 26-FEB-1988; 88US-00160771.
XX
PR 17-FEB-1988; 89US-00310881.
XX
PR 22-OCT-1990; 90US-00600244.
XX
PR 31-JUL-1992; 92US-00923692.
XX
PR 30-DEC-1992; 92US-00997733.
XX
PR 29-DEC-1993; 93US-00176414.
XX
PR 19-JAN-1994; 94US-00184237.
XX
PR 14-OCT-1994; 94US-00324003.
XX
PR 21-MAY-1999; 99US-00316572.
XX
PR 26-JUL-2000; 2000US-00636127.
XX
PR 13-NOV-2001; 2001US-00993059.
XX
(TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
PA (POGU/) POGUE G P.
PA (ERWT/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
PI
XX
XX WPI; 2004-142650/14.
XX
DR N-PSDB; ADM48682.
XX
XX New alpha-galactosidase polypeptides, useful in producing recombinant
PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
PT disease.
XX
XX Disclosure; SEQ ID NO 14; 72pp; English.
PS
XX
XX The present invention relates to novel galactosidase (Gal) proteins such
CC as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
CC of the present invention are useful for producing recombinant lysosomal
CC
```

```
enzymes for enzyme replacement therapy for treating human and animal
CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
CC human wild type rGAL-8R protein. This sequence comprises a human rGAL-8
CC protein and a ER retention signal. This sequence is used in the
CC exemplification of the invention.
XX
SQ Sequence 427 AA;
Query Match 100.0%; Score 2314; DB 8; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOLRNPELHLCALALRELALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MOLRNPELHLCALALRELALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
QY 61 DSCISEKLFMEWMAELMWSEGWKADAGEYELICIDDCWMAFQDSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEWMAELMWSEGWKADAGEYELICIDDCWMAFQDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSKGLGLGIYADVGNKTCAGPFGSGFYDYIDAQTFADWGVDLLKFDGICYDSLENL 180
Db 121 ANYVHSKGLGLGIYADVGNKTCAGPFGSGFYDYIDAQTFADWGVDLLKFDGICYDSLENL 180
QY 181 ADGYKHSALNRTGRSIVVSCWPLYNMFPQKPNYTEIROVCNHRNPFADIDDSWKSIIK 240
Db 181 ADGYKHSALNRTGRSIVVSCWPLYNMFPQKPNYTEIROVCNHRNPFADIDDSWKSIIK 240
QY 241 SILDWTSTFQRIIVDVAGPGWNDPMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFQRIIVDVAGPGWNDPMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
QY 361 GPRSYTIAVASIGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
Db 361 GPRSYTIAVASIGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 12
ADU66918
ID ADU66918 standard; protein; 427 AA.
XX
AC ADU66918;
XX
DT 10-FEB-2005 (first entry)
XX
DE Human alpha-galactosidase protein #6.
XX
KW Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human.
XX
OS Homo sapiens.
XX
PN US2004234516-A1.
XX
PD 25-NOV-2004.
XX
PF 21-MAY-2004; 2004US-00851388.
XX
PR 26-JUL-2000; 2000US-00626127.
XX
PR 13-NOV-2001; 2001US-00993059.
XX
PR 20-MAR-2002; 2002US-00103327.
XX
```

PA (LARG-) LARGE SCALE BIOLOGY CORP.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2004-821274/81.
XX N-PSDB; ADU66917.
XX A pharmaceutical composition comprising a lysosomal enzyme, useful for
PT enzyme replacement therapy for the treatment of lysosomal storage
PT diseases, such as Fabry's disease.
XX
XX
XX Disclosure; SEQ ID NO 14; 88pp; English.
XX
XX The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase enzymes having a post-translational modification provided
CC by the plant expression system. The invention is useful in enzyme
CC replacement therapy for treating lysosomal storage diseases such as
CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase protein.
XX
SQ Sequence 427 AA;
Query Match 100.0%; Score 2314; DB 8; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221; Indels 0; Gaps 0;
Matches 421; Conservative 0; Mismatches 0;
Qy 1 MQLNPHELHGCALALRFLALVSWDIPGARALDNGRLARTPTMGWLHWRFCNLDCCQSEP 60
Db 1 MQLNPHELHGCALALRFLALVSWDIPGARALDNGRLARTPTMGWLHWRFCNLDCCQSEP 60
Qy 61 DSCISEKLFMEAMELMWSEGWKDAYEYLCIDDCWMAQORDSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEAMELMWSEGWKDAYEYLCIDDCWMAQORDSEGRLOADPQRFPHGIRQL 120
Qy 121 ANYVHSKGLKLGIVADVGNKTCAGPFGSGFYVDIDAQTFADGWVLLKFDGCDLSLENL 180
Db 121 ANYVHSKGLKLGIVADVGNKTCAGPFGSGFYVDIDAQTFADGWVLLKFDGCDLSLENL 180
Qy 181 ADGYKMSLALNRGRSIVYSCWPLYMWPQKPNYTIROYCNHWRNFADIDDSWSKSIK 240
Db 181 ADGYKMSLALNRGRSIVYSCWPLYMWPQKPNYTIROYCNHWRNFADIDDSWSKSIK 240
Qy 241 SILDWTSPNQRIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSPNQRIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQSIG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQSIG 360
Qy 361 GPRSTYIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLSHINPTGTVLLOLENT 420
Db 361 GPRSTYIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLSHINPTGTVLLOLENT 420
Qy 421 M 421
Db 421 M 421
RESULT 13
ID AEA27447
XX AEA27447 standard; protein; 427 AA.
XX AEA27447;
XX AEA27447;
XX 11-AUG-2005 (first entry)
XX

DE Human alpha-galactosidase protein, rGAL-8R, SEQ ID NO: 14.
XX Gauchers disease; metabolic; neurological disease; niemann pick disease;
KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
KW antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
KW enzyme.
XX Homo sapiens.
OS
XX US2005125859-A1.
PN
XX 09-JUN-2005.
PD
XX 08-NOV-2004; 2004US-00984389.
PF
XX 26-JUL-2000; 2000US-00626127.
PR
XX 13-NOV-2001; 2001US-00993059.
PR
XX 20-MAR-2002; 2002US-00103327.
PR
XX (LARG-) LARGE SCALE BIOLOGY CORP.
PA Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2005-404004/41.
DR
XX N-PSDB; AEA27446.
DR
XX New isolated polypeptides useful for producing lysosomal enzymes in
PT plants to be utilized in enzyme replacement therapy or for the
PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
PT Gaucher's disease.
XX
XX Disclosure; SEQ ID NO 14; 88pp; English.
FS
XX The present invention relates to the production of human and animal
XX lysosomal enzymes in plants by a transient plant expression system. The
XX invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX galactosidase (gal) enzymes having a post-translational modification
XX provided by the plant expression system. The invention is useful in
XX enzyme replacement therapy for treating lysosomal storage diseases such
XX as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
XX disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX also useful in researches for developing new approaches to medical
XX treatment of lysosomal storage diseases and in industrial processes
XX involving enzymatic substrate hydrolysis. The present sequence is the
XX human alpha-galactosidase protein.
SQ Sequence 427 AA;
Query Match 100.0%; Score 2314; DB 9; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MQLNPHELHGCALALRFLALVSWDIPGARALDNGRLARTPTMGWLHWRFCNLDCCQSEP 60
Db 1 MQLNPHELHGCALALRFLALVSWDIPGARALDNGRLARTPTMGWLHWRFCNLDCCQSEP 60
Qy 61 DSCISEKLFMEAMELMWSEGWKDAYEYLCIDDCWMAQORDSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEAMELMWSEGWKDAYEYLCIDDCWMAQORDSEGRLOADPQRFPHGIRQL 120
Qy 121 ANYVHSKGLKLGIVADVGNKTCAGPFGSGFYVDIDAQTFADGWVLLKFDGCDLSLENL 180
Db 121 ANYVHSKGLKLGIVADVGNKTCAGPFGSGFYVDIDAQTFADGWVLLKFDGCDLSLENL 180
Qy 181 ADGYKMSLALNRGRSIVYSCWPLYMWPQKPNYTIROYCNHWRNFADIDDSWSKSIK 240
Db 181 ADGYKMSLALNRGRSIVYSCWPLYMWPQKPNYTIROYCNHWRNFADIDDSWSKSIK 240
Qy 241 SILDWTSPNQRIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSPNQRIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQSIG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQSIG 360

Db 301 RHISPOAKALLQDKVDVIAINODPLGKQGYQLRQGDNFVWEPRLSGLAWAVAMINRQBIG 360
 QY 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
 Db 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
 QY 421 M 421
 Db 421 M 421

RESULT 14

AAR53765

ID AAR53765 standard; protein; 429 AA.

XX AAR53765;

XX 25-MAR-2003 (revised)

DT 15-NOV-1994 (first entry)

XX Alpha-galactosidase A.

DE Alpha-galactosidase A.

XX Alpha-galactosidase A; Fabry disease; blood group O; blood group B; CHO;

KW Chinese hamster ovary; enzyme replacement therapy; glycoconjugate;

KW alpha-Gal A.

XX Homo sapiens.

OS Homo sapiens.

XX Key

FH Location/Qualifiers

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

FT Modified-site

FT /note= "putative N-glycosylation site"

Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MOLRNPHELGCALALREFALALVSWDIPGARALDNLGLARTPTMGWLHWRFPWNCNLDQCSEP 60
 Db 1 MOLRNPHELGCALALREFALALVSWDIPGARALDNLGLARTPTMGWLHWRFPWNCNLDQCSEP 60
 QY 61 DSCISEKLFMEWAEIWMVSEGKADAGEYELCIDDDCWMAFQDSEGRSQADPQRFPHGIRQL 120
 Db 61 DSCISEKLFMEWAEIWMVSEGKADAGEYELCIDDDCWMAFQDSEGRSQADPQRFPHGIRQL 120
 QY 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGYDYIDAQTFADWGVDDLKFDGICYCDLENL 180
 Db 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGYDYIDAQTFADWGVDDLKFDGICYCDLENL 180
 QY 181 ADGKXMSLALNRTGRSIVYSCWPLVMFPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
 Db 181 ADGKXMSLALNRTGRSIVYSCWPLVMFPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
 QY 241 SILDWTSEFQRIVDVAGPGWMDPDLVIGNFGLSMNQVQTQMALWAIMAAPLPMNDL 300
 Db 241 SILDWTSEFQRIVDVAGPGWMDPDLVIGNFGLSMNQVQTQMALWAIMAAPLPMNDL 300
 QY 301 RHISPOAKALLQDKVDVIAINODPLGKQGYQLRQGDNFVWEPRLSGLAWAVAMINRQBIG 360
 Db 301 RHISPOAKALLQDKVDVIAINODPLGKQGYQLRQGDNFVWEPRLSGLAWAVAMINRQBIG 360
 QY 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
 Db 361 GPRSYTIAVASLGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
 QY 421 M 421
 Db 421 M 421

RESULT 15

AAR63234

ID AAR63234 standard; protein; 429 AA.

XX AAR63234;

XX 29-JUN-1995 (first entry)

XX Human alpha-galactosidase A protein.

DE Human; alpha-galactosidase A; expression; mammalian; expression plasmid;

KW chromatography; fusion protein; Staphylococcus; protein A; cleavage site;

KW housekeeping gene; collagenase; affinity chromatography; galactosidase;

KW enzyme replacement therapy; gene therapy; lysosomal storage disorder;

KW Fabry disease; alpha-D-galacto-glycoconjugate; blood group.

XX Homo sapiens.

OS Homo sapiens.

XX Key

FH Location/Qualifiers

FT Modified-site

FT /note= "N-linked glycosylation"

FT Modified-site

FT /note= "N-linked glycosylation"

FT Modified-site

FT /note= "N-linked glycosylation"

FT Modified-site

FT /note= "N-linked glycosylation"

FT Modified-site

FT /note= "N-linked glycosylation"

FT Modified-site

FT /note= "N-linked glycosylation"

FT Modified-site

FT /note= "N-linked glycosylation"

FT Modified-site

FT /note= "N-linked glycosylation"

FT Modified-site

FT /note= "N-linked glycosylation"

FT Modified-site

FT /note= "N-linked glycosylation"

FT Modified-site

FT /note= "N-linked glycosylation"

FT Modified-site

FT /note= "N-linked glycosylation"

FT Modified-site

FT /note= "N-linked glycosylation"

Prod of human alpha-galactosidase A - by culturing cells contg the
 coding sequence and the beta-galactosidase alpha-2,6-sialyl transferase
 gene and regulation sequences.

Disclosure; Page 102-103; 156pp; English.

The cDNA sequence (AA066241) of human alpha-galactosidase A (alpha-Gal A)
 was determined. The predicted amino acid sequence (AAR53765) showed 50%
 homology with human alpha-N-acetylgalactosaminidase (AAR53766) (alpha-
 galactosidase B) and showed short regions of homology with yeast Mel 1
 (AAR53767). Recombinant alpha-Gal A was produced in CHO cells. The enzyme
 can be used for Fabry disease enzyme replacement therapy, to convert
 blood group B to O, or to hydrolyze alpha-D-galactosyl residues from
 glycoconjugates. (Updated on 25-MAR-2003 to correct PN field.)

Sequence 429 AA;

Query Match 100.0%; Score 2314; DB 2; Length 429;
 Best Local Similarity 100.0%; Pred. No. 1.8e-221;

(MOUN) MOUNT SINAI SCHOOL MEDICINE.

XX

PI Ioannou YA, Desnick RJ, Bishop DF;

XX WPI; 1994-340330/42.
DR N-PSDB; AAQ77861.

XX Recombinant human alpha-galactosidase A prodn. - using a mammalian host
PT cell expression system to obtain high yields of enzymatically active
PT enzyme.

XX Claim 3; Col 43-46; 60pp; English.

XX The amino acid sequence of the human alpha-galactosidase A (hAGA). The
CC gene encodes a protein of 429 amino acids. Methods to overexpress and
CC purify the protein include: (i) placing the 1.45 Kb human hAGA gene
CC fragment in a mammalian expression plasmid e.g. AAP91023 and purifying
CC total cell protein by chromatographic methods including affinity
CC chromatography or (ii) by producing a fusion protein comprising hAGA
CC fused to easily purified protein sequences e.g. the Staphylococcus protein
CC A, with cleavage site for a "housekeeping" gene e.g. collagenase cleavage
CC site, inserted between the two sequences in a mammalian expression
CC plasmid. The fusion protein can be easily purified by affinity
CC chromatography and the galactosidase protein released from the fusion
CC protein whilst still bound to the column by cleavage with collagenase.
CC The methods result in active alpha-galactosidase A protein. The hAGA can
CC be used for enzyme replacement therapy in patients with the lysosomal
CC storage disorder, Fabry disease. The protein can also be used in vitro to
CC modify alpha-D-galacto-glycoconjugates in a variety of processes e.g.
CC converting blood group B erythrocytes to group O

XX Sequence 429 AA;

Query Match 100.0%; Score 2314; DB 2; Length 429;

Best Local Similarity 100.0%; Pred. No. 1.8e-221;

Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MQLRNP	ELHLCALALR	FLALV	NDIP	GARALD	NGCLART	PTMG	WLHWR	FNC	NLDCQ	E	60
Db	1	MQLRNP	ELHLCALALR	FLALV	NDIP	GARALD	NGCLART	PTMG	WLHWR	FNC	NLDCQ	E	60
Qy	61	DSCISE	KLFMEAE	LMVSE	GWKD	AGYE	YLCID	DCW	MAPQ	RDS	EGRL	ADP	120
Db	61	DSCISE	KLFMEAE	LMVSE	GWKD	AGYE	YLCID	DCW	MAPQ	RDS	EGRL	ADP	120
Qy	121	ANYVH	SKGLGI	YAD	VGN	KTC	CAG	PGS	FGY	YD	IDA	QTF	180
Db	121	ANYVH	SKGLGI	YAD	VGN	KTC	CAG	PGS	FGY	YD	IDA	QTF	180
Qy	181	ADGYK	HMSL	ALNR	TGR	SI	VY	SC	EW	PL	YMW	PF	240
Db	181	ADGYK	HMSL	ALNR	TGR	SI	VY	SC	EW	PL	YMW	PF	240
Qy	241	SILD	WTSF	NQ	ERI	VD	V	AG	P	G	W	N	300
Db	241	SILD	WTSF	NQ	ERI	VD	V	AG	P	G	W	N	300
Qy	301	RHIS	PQ	AK	ALL	Q	D	K	D	V	I	A	360
Db	301	RHIS	PQ	AK	ALL	Q	D	K	D	V	I	A	360
Qy	361	GPR	SY	TI	AV	AS	L	G	K	G	V	A	420
Db	361	GPR	SY	TI	AV	AS	L	G	K	G	V	A	420
Qy	421	M	421										
Db	421	M	421										

Search completed: December 24, 2005, 08:54:18
Job time : 189 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 24, 2005, 08:54:29 ; Search time 166 Seconds
(without alignments)
1059.676 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MQLRNPGLHGCALALRFLA.....RLRSHINPTGVLLQLENTM 421

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA_Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	421	3	US-09-993-059-12
2	2314	100.0	421	4	US-10-103-327-12
3	2314	100.0	421	4	US-10-602-219-12
4	2314	100.0	421	4	US-10-602-220-12
5	2314	100.0	421	5	US-10-851-388-12
6	2314	100.0	421	5	US-10-851-389-12
7	2314	100.0	427	3	US-09-993-059-14
8	2314	100.0	427	4	US-10-103-327-14
9	2314	100.0	427	4	US-10-602-219-14
10	2314	100.0	427	4	US-10-602-220-14
11	2314	100.0	427	5	US-10-851-388-14
12	2314	100.0	427	5	US-10-851-389-14
13	2314	100.0	429	4	US-10-411-037-68
14	2314	100.0	429	4	US-10-411-026-68
15	2314	100.0	429	4	US-10-410-962-68
16	2314	100.0	429	4	US-10-411-049-68
17	2314	100.0	429	4	US-10-410-930-68
18	2314	100.0	429	4	US-10-410-997-68
19	2314	100.0	429	4	US-10-411-012-68
20	2314	100.0	429	4	US-10-410-913-68
21	2314	100.0	429	5	US-10-370-715B-278
22	2314	100.0	429	5	US-10-410-980-68
23	2314	100.0	429	5	US-10-410-897-68
24	2314	100.0	431	3	US-09-993-059-10
25	2314	100.0	431	4	US-10-103-327-10
26	2314	100.0	431	4	US-10-602-219-10
27	2314	100.0	431	4	US-10-602-220-10

28	2314	100.0	431	5	US-10-851-388-10	Sequence 10, Appl
29	2314	100.0	431	5	US-10-984-389-10	Sequence 10, Appl
30	2314	100.0	435	3	US-09-993-059-6	Sequence 6, Appl
31	2314	100.0	435	4	US-10-103-327-6	Sequence 6, Appl
32	2314	100.0	435	4	US-10-602-219-6	Sequence 6, Appl
33	2314	100.0	435	4	US-10-602-220-6	Sequence 6, Appl
34	2314	100.0	435	5	US-10-851-388-6	Sequence 6, Appl
35	2314	100.0	435	5	US-10-984-389-6	Sequence 6, Appl
36	2314	100.0	478	4	US-10-131-410-106	Sequence 106, App
37	2309	99.8	424	3	US-09-993-059-8	Sequence 8, Appl
38	2309	99.8	424	4	US-10-103-327-8	Sequence 8, Appl
39	2309	99.8	424	4	US-10-602-219-8	Sequence 8, Appl
40	2309	99.8	424	4	US-10-602-220-8	Sequence 8, Appl
41	2309	99.8	424	5	US-10-851-388-8	Sequence 8, Appl
42	2309	99.8	424	5	US-10-984-389-8	Sequence 8, Appl
43	2309	99.8	428	3	US-09-993-059-4	Sequence 4, Appl
44	2309	99.8	428	4	US-10-103-327-4	Sequence 4, Appl
45	2309	99.8	428	4	US-10-602-219-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-12
; Sequence 12, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-12

Query Match	100.0%	Score 2314;	DB 3;	Length 421;
Best Local Similarity	100.0%	Pred. No. 1.7e-221;		
Matches 421;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCMNLDCQBEP	60	
Db	1	MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCMNLDCQBEP	60	
Qy	61	DSCISEKLFMEAEIWMYSEGHKADAGEYLCITDDCWAPQDSEGLQADPQPFPHGIROL	120	
Db	61	DSCISEKLFMEAEIWMYSEGHKADAGEYLCITDDCWAPQDSEGLQADPQPFPHGIROL	120	
Qy	121	ANYVHSKGLKGIYADVGNKTCAGFPGSFGYYDIDAQTFADGWVDLLKFDGCGYCDLSLENL	180	
Db	121	ANYVHSKGLKGIYADVGNKTCAGFPGSFGYYDIDAQTFADGWVDLLKFDGCGYCDLSLENL	180	
Qy	181	ADGKHSMLALNRTGRSIVYCEWPLYNMPQKPNYTEIRQYCNHWRNFADIDDSNWSKTK	240	
Db	181	ADGKHSMLALNRTGRSIVYCEWPLYNMPQKPNYTEIRQYCNHWRNFADIDDSNWSKTK	240	
Qy	241	SILDWTSPNQRIVDVAGPGGNDPDLVIGNFGLSNNQOVTOMALWMAAPLPMNDL	300	
Db	241	SILDWTSPNQRIVDVAGPGGNDPDLVIGNFGLSNNQOVTOMALWMAAPLPMNDL	300	
Qy	301	RHISFOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWEPRLSGLAWAMINRQBIG	360	
Db	301	RHISFOAKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWEPRLSGLAWAMINRQBIG	360	
Qy	361	GPRSYTTAVASLGKGVACNPACFITQLLPVKKRLGFYEWTSRLRSHINPTGVLLQLENT	420	

Db 361 GPRSYTIAVSLGKGVACNPAFCFITQLLPVKKLGFEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 2

US-10-103-327-12
; Sequence 12, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-12

Query Match 100.0%; Score 2314; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.7e-221; Indels 0; Gaps 0;
Matches 421; Conservative 0; Mismatches 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWFRCMNCDCQEP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWFRCMNCDCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAFORDSEGRLOADPQRFPHGIROL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAFORDSEGRLOADPQRFPHGIROL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGYDYIDAQTFADWGVLLKFDGVCDSLENL 180
Db 121 ANYVHSKGLGIYADVGNKTCAGPFGSGYDYIDAQTFADWGVLLKFDGVCDSLENL 180
QY 181 ADGKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROQCNHWRNFADIDDSWKSJK 240
Db 181 ADGKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROQCNHWRNFADIDDSWKSJK 240
QY 241 SILDWTSTFQNERIVDVAGPGGWNDDPMLVIGNFGLSWNQVVTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFQNERIVDVAGPGGWNDDPMLVIGNFGLSWNQVVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVSLGKGVACNPAFCFITQLLPVKKLGFEWTSRLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTIAVSLGKGVACNPAFCFITQLLPVKKLGFEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 3

US-10-602-219-12
; Sequence 12, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:

; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-602-219-12

Query Match 100.0%; Score 2314; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.7e-221; Indels 0; Gaps 0;
Matches 421; Conservative 0; Mismatches 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWFRCMNCDCQEP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWFRCMNCDCQEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAFORDSEGRLOADPQRFPHGIROL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAFORDSEGRLOADPQRFPHGIROL 120
QY 121 ANYVHSKGLGIYADVGNKTCAGPFGSGYDYIDAQTFADWGVLLKFDGVCDSLENL 180
Db 121 ANYVHSKGLGIYADVGNKTCAGPFGSGYDYIDAQTFADWGVLLKFDGVCDSLENL 180
QY 181 ADGKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROQCNHWRNFADIDDSWKSJK 240
Db 181 ADGKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROQCNHWRNFADIDDSWKSJK 240
QY 241 SILDWTSTFQNERIVDVAGPGGWNDDPMLVIGNFGLSWNQVVTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFQNERIVDVAGPGGWNDDPMLVIGNFGLSWNQVVTQMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQETG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQETG 360
QY 361 GPRSYTIAVSLGKGVACNPAFCFITQLLPVKKLGFEWTSRLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTIAVSLGKGVACNPAFCFITQLLPVKKLGFEWTSRLRSHINPTGTVLLQLENT 420
QY 421 M 421
Db 421 M 421

RESULT 4
US-10-602-220-12
; Sequence 12, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-602-220-12

Query Match 100.0%; Score 2314; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.7e-221; Mismatches 0; Indels 0; Gaps 0;
Matches 421; Conservative 0;
Qy 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
Db 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAFORDSEGRLOADPQRFPHGIROL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAFORDSEGRLOADPQRFPHGIROL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGPFGSGFYDDIDAOQTFADGWVLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGPFGSGFYDDIDAOQTFADGWVLLKFDGCGYCDLSLENL 180
Qy 181 ADGKHMSLALNRTGRSIVYSCWPLYNMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKHMSLALNRTGRSIVYSCWPLYNMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIINODPLGKQGYQLRQGNFVFWERPLSLGLAWAVAMINRQIEG 360
Db 301 RHISPOAKALLQDKDVIINODPLGKQGYQLRQGNFVFWERPLSLGLAWAVAMINRQIEG 360

Qy 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Qy 421 M 421
Db 421 M 421
RESULT 5
US-10-851-388-12
; Sequence 12, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: TURPEN, Stephen A.
; APPLICANT: GARGER, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-851-388-12

Query Match 100.0%; Score 2314; DB 5; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.7e-221; Mismatches 0; Indels 0; Gaps 0;
Matches 421; Conservative 0;
Qy 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
Db 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAFORDSEGRLOADPQRFPHGIROL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAFORDSEGRLOADPQRFPHGIROL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGPFGSGFYDDIDAOQTFADGWVLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGPFGSGFYDDIDAOQTFADGWVLLKFDGCGYCDLSLENL 180
Qy 181 ADGKHMSLALNRTGRSIVYSCWPLYNMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKHMSLALNRTGRSIVYSCWPLYNMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIINODPLGKQGYQLRQGNFVFWERPLSLGLAWAVAMINRQIEG 360
Db 301 RHISPOAKALLQDKDVIINODPLGKQGYQLRQGNFVFWERPLSLGLAWAVAMINRQIEG 360
Qy 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Qy 421 M 421
Db 421 M 421

RESULT 6
US-10-984-389-12
; Sequence 12, Application US/10984389

```
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-984-389-12

Query Match      100.0%; Score 2314; DB 5; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.7e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60
DB      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60
QY      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120
DB      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120
QY      121  ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180
DB      121  ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180
QY      181  ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSITK 240
DB      181  ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSITK 240
QY      241  SILDWTSTFNOERIVDVAGPGWNPDMVLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
DB      241  SILDWTSTFNOERIVDVAGPGWNPDMVLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
QY      301  RHISPOAKALLQDKDVIANQDPLGKQGYQLRQGNFVWERPLSLGLAWAVAMINRQETG 360
DB      301  RHISPOAKALLQDKDVIANQDPLGKQGYQLRQGNFVWERPLSLGLAWAVAMINRQETG 360
QY      361  GPRSTYIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB      361  GPRSTYIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY      421  M 421
DB      421  M 421

RESULT 7
US-09-993-059-14
; Sequence 14, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT FILING DATE: 2004-11-13
; PRIOR APPLICATION NUMBER: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-984-389-12

Query Match      100.0%; Score 2314; DB 5; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.7e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60
DB      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60
QY      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120
DB      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120
QY      121  ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180
DB      121  ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180
QY      181  ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSITK 240
DB      181  ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSITK 240
QY      241  SILDWTSTFNOERIVDVAGPGWNPDMVLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
DB      241  SILDWTSTFNOERIVDVAGPGWNPDMVLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
QY      301  RHISPOAKALLQDKDVIANQDPLGKQGYQLRQGNFVWERPLSLGLAWAVAMINRQETG 360
DB      301  RHISPOAKALLQDKDVIANQDPLGKQGYQLRQGNFVWERPLSLGLAWAVAMINRQETG 360
QY      361  GPRSTYIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB      361  GPRSTYIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY      421  M 421
DB      421  M 421

RESULT 8
US-10-103-327-14
; Sequence 14, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-14

Query Match      100.0%; Score 2314; DB 4; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60
DB      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60
QY      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120
DB      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120
QY      121  ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180
DB      121  ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180
QY      181  ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSITK 240
DB      181  ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSITK 240
QY      241  SILDWTSTFNOERIVDVAGPGWNPDMVLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
DB      241  SILDWTSTFNOERIVDVAGPGWNPDMVLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
QY      301  RHISPOAKALLQDKDVIANQDPLGKQGYQLRQGNFVWERPLSLGLAWAVAMINRQETG 360
DB      301  RHISPOAKALLQDKDVIANQDPLGKQGYQLRQGNFVWERPLSLGLAWAVAMINRQETG 360
QY      361  GPRSTYIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB      361  GPRSTYIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY      421  M 421
DB      421  M 421
```

```
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-14

Query Match      100.0%; Score 2314; DB 3; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60
DB      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60
QY      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120
DB      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120
QY      121  ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180
DB      121  ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180
QY      181  ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSITK 240
DB      181  ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSITK 240
QY      241  SILDWTSTFNOERIVDVAGPGWNPDMVLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
DB      241  SILDWTSTFNOERIVDVAGPGWNPDMVLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
QY      301  RHISPOAKALLQDKDVIANQDPLGKQGYQLRQGNFVWERPLSLGLAWAVAMINRQETG 360
DB      301  RHISPOAKALLQDKDVIANQDPLGKQGYQLRQGNFVWERPLSLGLAWAVAMINRQETG 360
QY      361  GPRSTYIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB      361  GPRSTYIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY      421  M 421
DB      421  M 421

RESULT 8
US-10-103-327-14
; Sequence 14, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-14

Query Match      100.0%; Score 2314; DB 4; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60
DB      1  MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCMNLDCQEEP 60
QY      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120
DB      61  DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAPORDSEGRLOADPQRFPHGIROL 120
QY      121  ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180
DB      121  ANYVHSGKGLGIYADVGNKTCAGPGSGFYVDIDAQTFADWGVLLKFDGCGYCDLSLENL 180
QY      181  ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSITK 240
DB      181  ADGYKXMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIROYCNHWRNFADIDDSWKSITK 240
QY      241  SILDWTSTFNOERIVDVAGPGWNPDMVLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
DB      241  SILDWTSTFNOERIVDVAGPGWNPDMVLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
QY      301  RHISPOAKALLQDKDVIANQDPLGKQGYQLRQGNFVWERPLSLGLAWAVAMINRQETG 360
DB      301  RHISPOAKALLQDKDVIANQDPLGKQGYQLRQGNFVWERPLSLGLAWAVAMINRQETG 360
QY      361  GPRSTYIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
DB      361  GPRSTYIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
QY      421  M 421
DB      421  M 421
```

[illegible]

RESULT 9
 US-10-602-219-14
 ; Sequence 14, Application US/10602219
 ; Publication No. US20040016021A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Large Scale Biology Corporation
 ; APPLICANT: Turpen, Thomas H.
 ; APPLICANT: Pogue, Gregory P.
 ; APPLICANT: Erwin, Robert L.
 ; APPLICANT: Grill, Laurence K.
 ; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
 ; FILE REFERENCE: L8BC-0087-CP09B
 ; CURRENT APPLICATION NUMBER: US/10/602,219
 ; CURRENT FILING DATE: 2003-05-23
 ; PRIOR APPLICATION NUMBER: 09/993,059
 ; PRIOR FILING DATE: 2001-11-13
 ; PRIOR APPLICATION NUMBER: 09/626,127
 ; PRIOR FILING DATE: 2000-07-26
 ; PRIOR APPLICATION NUMBER: 09/316,572
 ; PRIOR FILING DATE: 1999-05-21
 ; PRIOR APPLICATION NUMBER: 08/324,003
 ; PRIOR FILING DATE: 1994-10-14
 ; PRIOR APPLICATION NUMBER: 08/176,414
 ; PRIOR FILING DATE: 1993-12-29
 ; PRIOR APPLICATION NUMBER: 07/997,733
 ; PRIOR FILING DATE: 1992-12-30
 ; PRIOR APPLICATION NUMBER: 08/184,237
 ; PRIOR FILING DATE: 1994-01-19
 ; PRIOR APPLICATION NUMBER: 07/923,692
 ; PRIOR FILING DATE: 1992-07-31
 ; PRIOR APPLICATION NUMBER: 07/600,244
 ; PRIOR FILING DATE: 1990-10-22
 ; PRIOR APPLICATION NUMBER: 07/641,617
 ; PRIOR FILING DATE: 1991-01-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 32
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 14
 ; LENGTH: 427
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-602-219-14

Query Match

[illegible]

```

RESULT 10
US-10-602-220-14
; Sequence 14, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32

```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-602-220-14

Query Match      100.0%; Score 2314; DB 4; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQSEP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQSEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGPGSGFYGYDIDAQTFADWGVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKGLGIYADVGNKTCAGPGSGFYGYDIDAQTFADWGVDLLKFDGCGYCDLSLENL 180
QY 181 ADGYKXMSLALNRTGRSIVYSCWPLVYMWPPQKNYTEIROQCNHWRNFADIDDSWKSIIK 240
Db 181 ADGYKXMSLALNRTGRSIVYSCWPLVYMWPPQKNYTEIROQCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOQVTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOQVTOMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQBIG 360
Db 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQBIG 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
QY 421 M 421
Db 421 M 421

RESULT 12
US-10-984-389-14
; Sequence 14, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-984-389-14

Query Match      100.0%; Score 2314; DB 5; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQSEP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQSEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGPGSGFYGYDIDAQTFADWGVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKGLGIYADVGNKTCAGPGSGFYGYDIDAQTFADWGVDLLKFDGCGYCDLSLENL 180
QY 181 ADGYKXMSLALNRTGRSIVYSCWPLVYMWPPQKNYTEIROQCNHWRNFADIDDSWKSIIK 240
Db 181 ADGYKXMSLALNRTGRSIVYSCWPLVYMWPPQKNYTEIROQCNHWRNFADIDDSWKSIIK 240
Db 181 ADGYKXMSLALNRTGRSIVYSCWPLVYMWPPQKNYTEIROQCNHWRNFADIDDSWKSIIK 240

Query Match      100.0%; Score 2314; DB 4; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQSEP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQSEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYELCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGPGSGFYGYDIDAQTFADWGVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKGLGIYADVGNKTCAGPGSGFYGYDIDAQTFADWGVDLLKFDGCGYCDLSLENL 180
QY 181 ADGYKXMSLALNRTGRSIVYSCWPLVYMWPPQKNYTEIROQCNHWRNFADIDDSWKSIIK 240
Db 181 ADGYKXMSLALNRTGRSIVYSCWPLVYMWPPQKNYTEIROQCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOQVTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFNOERIVDVAGPGGWNDDMLVIGNFGLSNQOQVTOMALWAIMAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQBIG 360
Db 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVFWERPLSGLAWAVAMINRQBIG 360
QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOLENT 420
QY 421 M 421
Db 421 M 421

RESULT 11
US-10-851-388-14
; Sequence 14, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-851-388-14

Query Match      100.0%; Score 2314; DB 5; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNCDCQSEP 60
```

Qy 241 SILDWTSPNQRIRVDVAGPGGNDPDMVLVGNFGLSNWQVQTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSPNQRIRVDVAGPGGNDPDMVLVGNFGLSNWQVQTOMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
Db 301 RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
Qy 361 GPRSYTTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Qy 421 M 421
Db 421 M 421

RESULT 13

US-10-411-037-68
; Sequence 68, Application US/10411037
; Publication No. US20040043446A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: DeFrees, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; APPLICANT: Bove, Caryn
; TITLE OF INVENTION: ALPHA GALACTOSIDASE A: REMODELING AND GLYCOCONJUGATION OF ALPHA
; FILE REFERENCE: 040853-01-5082
; CURRENT APPLICATION NUMBER: US/10/411,037
; CURRENT FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; PRIOR FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-411-037-68

Query Match 100.0%; Score 2314; DB 4; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.8e-221; Indels 0; Gaps 0;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
Db 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWDAGYEYLCIDDCWMAFQDSRGLQADPQRFPHGIQRL 120
Db 61 DSCISEKLFMEAEMLVSEGWDAGYEYLCIDDCWMAFQDSRGLQADPQRFPHGIQRL 120
Qy 121 ANYHSGKLGKIYADVGNKTCAGPFGSGFYDDAQTFFADGWVDLLKFDGCYCDLSNLL 180
Db 121 ANYHSGKLGKIYADVGNKTCAGPFGSGFYDDAQTFFADGWVDLLKFDGCYCDLSNLL 180

Qy 181 ADGKXKMSLALNRTGRSIVYSCWPLWMPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKXKMSLALNRTGRSIVYSCWPLWMPQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSPNQRIRVDVAGPGGNDPDMVLVGNFGLSNWQVQTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSPNQRIRVDVAGPGGNDPDMVLVGNFGLSNWQVQTOMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
Db 301 RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQBIG 360
Qy 361 GPRSYTTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Db 361 GPRSYTTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQLENT 420
Qy 421 M 421
Db 421 M 421

RESULT 14

US-10-411-026-68
; Sequence 68, Application US/10411026
; Publication No. US20040063911A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: DeFrees, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; TITLE OF INVENTION: PROTEIN REMODELING METHODS AND PROTEINS/PEPTIDES PRODUCED BY THE
; FILE REFERENCE: 040853-01-5053
; CURRENT APPLICATION NUMBER: US/10/411,026
; CURRENT FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; PRIOR FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-411-026-68

Query Match 100.0%; Score 2314; DB 4; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.8e-221; Indels 0; Gaps 0;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
Db 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWDAGYEYLCIDDCWMAFQDSRGLQADPQRFPHGIQRL 120
Db 61 DSCISEKLFMEAEMLVSEGWDAGYEYLCIDDCWMAFQDSRGLQADPQRFPHGIQRL 120
Qy 121 ANYHSGKLGKIYADVGNKTCAGPFGSGFYDDAQTFFADGWVDLLKFDGCYCDLSNLL 180

Db 121 ANYVHSKGLKGIYADVGNKTCAGFPGSFGYYDIDAQTFADWGVLLKFDGVCDSLENL 180
Qy 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSTFNOERIIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFNOERIIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQBIG 360
Db 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQBIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLOLENT 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLOLENT 420
Qy 421 M 421
Db 421 M 421

RESULT 15

US-10-410-962-68
; Sequence 68, Application US/10410962
; Publication No. US20040077836A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: Defrees, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; APPLICANT: Bowe, Caryn
; TITLE OF INVENTION: GLYCOCONJUGATION OF G-CSF
; FILE REFERENCE: 040853-01-5054
; CURRENT APPLICATION NUMBER: US/10/410,962
; PRIOR FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; PRIOR FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-410-962-68

Query Match 100.0%; Score 2314; DB 4; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.8e-221;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MQLRNPDLHGCALALRFLALVSWDIPGARALDNLGARTPTMGWLHWRFCNLDQCEEP 60
Db 1 MQLRNPDLHGCALALRFLALVSWDIPGARALDNLGARTPTMGWLHWRFCNLDQCEEP 60
Qy 61 DSCISEKLFMEAMLMVSEGWKDGAGYEYLCIDDCWMAPOQDSEGRLOADPQRPFGIROL 120
Db 61 DSCISEKLFMEAMLMVSEGWKDGAGYEYLCIDDCWMAPOQDSEGRLOADPQRPFGIROL 120

Qy 121 ANYVHSKGLKGIYADVGNKTCAGFPGSFGYYDIDAQTFADWGVLLKFDGVCDSLENL 180
Db 121 ANYVHSKGLKGIYADVGNKTCAGFPGSFGYYDIDAQTFADWGVLLKFDGVCDSLENL 180
Qy 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSTFNOERIIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSTFNOERIIVDVAGPGGWNDDMLVIGNFGLSNQOVTOMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQBIG 360
Db 301 RHISPOAKALLQDKDVIAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINRQBIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLOLENT 420
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLOLENT 420
Qy 421 M 421
Db 421 M 421

Search completed: December 24, 2005, 09:11:10
Job time : 168 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 24, 2005, 08:50:53 ; Search time 13 Seconds
(without alignments)
230.993 Million cell updates/sec

Title: us-10-602-219-12

Perfect score: 2314

Sequence: 1 MQLRNPGLHLCALALRFLA.....RLRSHINPTGVLLQLENTM 421

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 54001 seqs, 7132810 residues

Total number of hits satisfying chosen parameters: 54001

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA New:

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/PC7_NEW_PUB pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85	3.7	437	7	US-11-073-626-3
2	82	3.5	437	7	US-11-073-626-1
3	81	3.5	310	7	US-11-074-176-108
4	81	3.5	944	7	US-11-057-058-68
5	80.5	3.5	414	6	US-10-467-657-32
6	80.5	3.5	414	6	US-10-467-657-8372
7	80	3.5	861	7	US-11-038-284-36
8	79.5	3.4	1827	7	US-11-057-058-62
9	79	3.4	322	6	US-10-467-657-8554
10	79	3.4	368	7	US-11-129-143-62
11	78.5	3.4	622	6	US-10-510-386-236
12	78	3.4	760	6	US-10-821-234-1141
13	78	3.4	786	6	US-10-467-962B-103
14	77.5	3.3	522	6	US-10-131-826A-450
15	77.5	3.3	903	7	US-11-057-058-65
16	77	3.3	734	6	US-10-652-893-2
17	76.5	3.3	320	6	US-10-467-657-3254
18	76.5	3.3	717	6	US-10-793-626-3022
19	76	3.3	2769	7	US-11-113-424-14
20	75.5	3.3	275	7	US-11-055-822-2
21	75.5	3.3	843	6	US-10-645-441-8
22	75	3.2	402	6	US-10-650-326B-21
23	75	3.2	402	6	US-11-051-568-29
24	75	3.2	627	6	US-10-873-528-191
25	74.5	3.2	482	6	US-10-793-626-1314

26 74.5 3.2 573 6 US-10-467-657-684 Sequence 684, Appl
27 74.5 3.2 966 7 US-11-057-058-67 Sequence 67, Appl
28 74 417 6 US-10-467-657-6612 Sequence 6612, Ap
29 73.5 3.2 932 7 US-11-057-058-59 Sequence 59, Appl
30 73 3.2 514 7 US-11-103-037-3 Sequence 3, Appl
31 73 3.2 734 7 US-11-137-455-65 Sequence 65, Appl
32 72 3.1 515 6 US-10-630-203-6 Sequence 6, Appl
33 72 3.1 842 7 US-11-038-284-38 Sequence 38, Appl
34 72 3.1 847 7 US-11-038-284-42 Sequence 42, Appl
35 70.5 3.0 316 6 US-10-821-234-1526 Sequence 1526, Ap
36 70.5 3.0 834 6 US-10-928-446A-188 Sequence 188, App
37 70.5 3.0 834 6 US-10-928-446A-192 Sequence 192, App
38 70.5 3.0 834 6 US-10-928-446A-194 Sequence 194, App
39 70.5 3.0 834 6 US-10-928-446A-196 Sequence 196, App
40 70.5 3.0 834 6 US-10-928-446A-198 Sequence 198, App
41 70.5 3.0 834 6 US-10-928-446A-200 Sequence 200, App
42 70.5 3.0 834 6 US-10-928-446A-202 Sequence 202, App
43 70.5 3.0 942 6 US-10-928-446A-190 Sequence 190, App
44 70.5 3.0 947 6 US-10-928-446A-182 Sequence 182, App
45 70.5 3.0 950 6 US-10-928-446A-184 Sequence 184, App

ALIGNMENTS

RESULT 1

US-11-073-626-3
; Sequence 3, Application US/11073626
; Publication No. US20050244926A1
; GENERAL INFORMATION:

; APPLICANT: KUROSAWA, KEIKO
; APPLICANT: HIROKAWA, KOZO
; APPLICANT: KAJIYAMA, NAOKI
; TITLE OF INVENTION: Novel fructosyl peptide oxidase
; FILE REFERENCE: 227590US0
; CURRENT APPLICATION NUMBER: US/11/073,626
; CURRENT FILING DATE: 2005-03-08
; PRIOR APPLICATION NUMBER: US/10/232,655
; PRIOR FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: JP 2001-266665
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: JP 2001-378151
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: JP 2002-228727
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Eupenicillium terrenum
US-11-073-626-3

Query Match 3.7%; Score 85; DB 7; Length 437;

Best Local Similarity 20.4%; Pred. No. 0.62;
Matches 46; Conservative 29; Mismatches 71; Indels 80; Gaps 12;

Qy 32 LDNGLRPTMGWLHWRFCNLDQCPEPDCISEKLFMEAEMLVMSGWDAGYEYLCI 91
Db 115 LDAGIGLEKTNVWL-----ESEDILAKAPNFTREQV---KGNKG-----LPCT 155
Qy 92 DDCWNPORDSEGRLOADPQRPFHGIROLANYVHSGKGLKG-----IYADVGNKT 141
Db 156 DGGWLAARK-----AINAIGIFLDQKGVKFGGAGTQQPLFAADG-KT 199
Qy 142 CAGPFGSGYDYIDAQTFAD-----WGVDLKF-DGC-----YCDSELENADGY 184
Db 200 CIGLETTD-----TKVPADKVLAAAGAWSPTLDLEDQCVSKAWVFAHIQUTPKEDAY 254
Qy 185 KHMSLALNRTGRSIVYSCENPLYMPPFOKPNVTEIROYCNHWRNFA 230
Db 255 KNV-----PVYDGEYGF-----FFEPNEYGVIKVCDEFFPGFS 287

RESULT 2

US-11-073-626-1
; Sequence 1, Application US/11073626
; Publication No. US20050244926A1
; GENERAL INFORMATION:
; APPLICANT: HIROKAWA, KEIKO
; APPLICANT: KUROKAWA, KOZO
; APPLICANT: KAJIYAMA, NAOKI
; TITLE OF INVENTION: Novel fructosyl peptide oxidase
; FILE REFERENCE: 227590U0
; CURRENT APPLICATION NUMBER: US/11/073,626
; CURRENT FILING DATE: 2005-03-08
; PRIOR APPLICATION NUMBER: US/10/232,655
; PRIOR FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: JP 2001-266665
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: JP 2001-378151
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: JP 2002-228727
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Coniochaeta sp.
US-11-073-626-1

Query Match 3.5%; Score 82; DB 7; Length 437;
Best Local Similarity 19.5%; Pred. No. 1.2;
Matches 42; Conservative 31; Mismatches 82; Indels 60; Gaps 10;
QY 33 DNLGARTPMGLHWRFCNLDQBEPPDSCISEKLFMEAEMLVSEGWKAGYELCID 92
DB 116 DAGAGLEKTHAWL-----DNEDEILSKMPLIQDQI---QGWKAIWSQ---D 156
QY 93 DCWMAQRDSEGLQADPQRPFGHRLQRLANVHSGKLGK-----IYADVGNKTC 142
DB 157 GGLAAAK-----AATGQFLKRGVFGGAGSFQKPLFDDGET-TC 200
QY 143 AGFPGSFG---YYIDAQTFADWGVLLKFDGCYCDLSLENLADGYKHMSL-----ALNRTG 195
DB 201 IGVETADGPKYADKVLVAAGAWSPTLVDLQDCCSK---AWVYAHQLTPEEAAYKG 256
QY 196 RSIVVSCWPLVWFPQKPNYTEIRQYCNHWRNFA 230
DB 257 VPVVYNGEFGF----FFEFNEFGVIKVCDEFPFGS 287

RESULT 3

US-11-074-176-108
; Sequence 108, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klienhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 108
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus

US-11-074-176-108

Query Match 3.5%; Score 81; DB 7; Length 310;
Best Local Similarity 20.9%; Pred. No. 0.98;
Matches 54; Conservative 32; Mismatches 114; Indels 58; Gaps 10;
QY 56 QCEPDCISBKLPMEMAEMLVSEGWKADAGYELCIDDCWMAQRDSEGLQADPQRPFH 115
DB 10 CYNEEESI---PLFYNAVB-KVMQTILDLDEYFWFVNDSSDKSLEENRQKQKNDPEHVH 65
QY 116 GROLANVHSGKLGKGIYADVGNKTCAGFPGSGYGYDIDAQTFADWGVLLKFDGCYCD 175
DB 66 YVFSRNFGEAAALYAGLQASIGDYIVV-----MDVDLQ-----D 100
QY 176 SLENLADGYKMSLA-----LNRGTSIVVSCWPLVWFPQKPNYTEI---ROY 222
DB 101 PPEFLPEMYNIKTGYDCICTRRVDRKGEAKFKSFLSDMFYKLVNKSINTEIVSGARDY 160
QY 223 CNHWRNFAD---IDDSWKSISILDWTSF-----NQRIVDVAGPGGWNDPDMLEVI 270
DB 161 RMTROMVDVAVLSMPEYNRFSKGFISWGVFKTKYLDYHNVH---VAGESDNTWKLFKY 217
QY 271 GNFGLS-WNQVQTQMALW 287
DB 218 AMDGIADFSQAPLNIAVW 235

RESULT 4

US-11-057-058-68
; Sequence 68, Application US/11057058
; Publication No. US2005024400A1
; GENERAL INFORMATION:
; APPLICANT: Lebowitz, Jonathan
; APPLICANT: Maga, John
; TITLE OF INVENTION: ACID ALPHA-GLUCOSIDASE AND FRAGMENTS THEREOF
; FILE REFERENCE: SYM-011
; CURRENT APPLICATION NUMBER: US/11/057,058
; CURRENT FILING DATE: 2005-02-10
; PRIOR APPLICATION NUMBER: US 60/543,812
; PRIOR FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 68
; SEQ ID NO 68
; LENGTH: 944
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-057-058-68

Query Match 3.5%; Score 81; DB 7; Length 944;
Best Local Similarity 26.5%; Pred. No. 4.3;
Matches 36; Conservative 13; Mismatches 49; Indels 38; Gaps 7;
QY 75 LMVSEGMKADAGYELCIDDCWMAQRDSEGLQADPQRPFGH-----IFQLAN----- 122
DB 410 LEVDQGFDD---HNLPCDVIWLDIEHADGKRYFTWDPSPFPQPTMLERLASKRRKLVAI 466
QY 123 -----YVHSGKLGKGIYADVGNKTCAG-----PQSGYGYDIDAQTFADWGV 164
DB 467 VDPHIKVDGYSYRVHLEENLGLYV---KTRDGSYEGWCWPGSAGYPDFNTPTWAWWA 522
QY 165 DLLKFDGCYCDLSLENI 180
DB 523 NMFSDYN-YEGSAPNL 537

RESULT 5

US-10-467-657-32
; Sequence 32, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia

APPLICANT: MASIGNANI Vega
APPLICANT: MONACI Elisabetta
TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/467,657
CURRENT FILING DATE: 2003-08-11
PRIOR APPLICATION NUMBER: GB-0103424.8
PRIOR FILING DATE: 2001-02-12
NUMBER OF SEQ ID NOS: 9218
SOFTWARE: SeqWin99, version 1.04
SEQ ID NO 32
LENGTH: 414
TYPE: PRT
ORGANISM: Neisseria gonorrhoeae
US-10-467-657-32

Query Match 3.5%; Score 80.5; DB 6; Length 414;
Best Local Similarity 19.4%; Pred. No. 1.6;
Matches 37; Conservative 30; Mismatches 79; Indels 45; Gaps 7;
Qy 17 RFLALVWDIPGARALDNGLARTPTMGWLHWRFCNLDCEPDSCISEKLFMEMAELM 76
Db 264 RFREIGNRLAELNRIGNAFEDTPI-----EVLILLVDKDKTDGGCIRELYDCKTDTLL 317
Qy 77 VSEGWKAGYEYLCIDCCWMAPORESEGRLOADPQPFPHGIROLANYVHSGKLGIYAD 136
Db 318 AADTWQ-----IEPDLWQTVEPAPPKKEDPVLLEHRCRDAAKRIARELR---FSK 367
Qy 137 VGNKTCAGFGSGFYDIDAQTFADMGVDLLKFGCCYCDLSLENLADGKHKMSLALNRTGR 196
Db 368 MVNE-IEGWPHA-----EFDG-FCDLRLCNLIQAKEY-----GK 398
Qy 197 SIVVSCWPLY 207
Db 399 KHYFPCSLPLF 409

RESULT 6
US-10-467-657-8372
Sequence 8372, Application US/10467657
Publication No. US20050260581A1
GENERAL INFORMATION:
APPLICANT: CHIRON SpA
APPLICANT: FONTANA Maria Rita
APPLICANT: PIZZA Mariagrazia
APPLICANT: MASIGNANI Vega
APPLICANT: MONACI Elisabetta
TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/467,657
CURRENT FILING DATE: 2003-08-11
PRIOR APPLICATION NUMBER: GB-0103424.8
PRIOR FILING DATE: 2001-02-12
NUMBER OF SEQ ID NOS: 9218
SOFTWARE: SeqWin99, version 1.04
SEQ ID NO 8372
LENGTH: 414
TYPE: PRT
ORGANISM: Neisseria gonorrhoeae
US-10-467-657-8372

Query Match 3.5%; Score 80.5; DB 6; Length 414;
Best Local Similarity 19.4%; Pred. No. 1.6;
Matches 37; Conservative 30; Mismatches 79; Indels 45; Gaps 7;
Qy 17 RFLALVWDIPGARALDNGLARTPTMGWLHWRFCNLDCEPDSCISEKLFMEMAELM 76
Db 264 RFREIGNRLAELNRIGNAFEDTPI-----EVLILLVDKDKTDGGCIRELYDCKTDTLL 317
Qy 77 VSEGWKAGYEYLCIDCCWMAPORESEGRLOADPQPFPHGIROLANYVHSGKLGIYAD 136
Db 318 AADTWQ-----IEPDLWQTVEPAPPKKEDPVLLEHRCRDAAKRIARELR---FSK 367

Qy 137 VGNKTCAGFGSGFYDIDAQTFADMGVDLLKFGCCYCDLSLENLADGKHKMSLALNRTGR 196
Db 368 MVNE-IEGWPHA-----EFDG-FCDLRLCNLIQAKEY-----GK 398
Qy 197 SIVVSCWPLY 207
Db 399 KHYFPCSLPLF 409
RESULT 7
US-11-038-284-36
Sequence 36, Application US/11038284
Publication No. US20050246793A1
GENERAL INFORMATION:
APPLICANT: COOKE, DAVID
APPLICANT: DEBET, MARTINE
APPLICANT: GIDLEY, MICHAEL, JOHN
APPLICANT: JOBLING, STEPHEN, ALAN
APPLICANT: SAFFORD, RICHARD
APPLICANT: SIDEBOTTOM, CHRISTOPHER, MICHAEL
APPLICANT: WESTCOTT, ROGER, JOHN
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO PLANT STARCH COMPOSITION
FILE REFERENCE: 054163-5003-US
CURRENT APPLICATION NUMBER: US/11/038,284
CURRENT FILING DATE: 2005-01-21
PRIOR APPLICATION NUMBER: US/10/056,454
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: PCT/GB96/01075
PRIOR FILING DATE: 1996-05-03
PRIOR APPLICATION NUMBER: GB 9607409.1
PRIOR FILING DATE: 1996-04-10
PRIOR APPLICATION NUMBER: GB 9509229.2
PRIOR FILING DATE: 1995-05-05
NUMBER OF SEQ ID NOS: 43
SOFTWARE: PatentIn version 3.2
SEQ ID NO 36
LENGTH: 861
TYPE: PRT
ORGANISM: Lathyrus sp.
US-11-038-284-36

Query Match 3.5%; Score 80; DB 7; Length 861;
Best Local Similarity 25.6%; Pred. No. 4.8;
Matches 32; Conservative 17; Mismatches 42; Indels 34; Gaps 6;
Qy 108 ADPQRF--PHGIROLANYVHSGKLK---GIYADVGNKTCAGF-----PGSFG 150
Db 395 APSRFGTPELKSILDRHELGLLVLMDIYVHSSNNVTLGLNNFDTGDGHYFPHGSRG 454
Qy 151 YYDI-DAQTF-----ADMGVDLLKFGCCYCDLSLENLADGKHKMSLALNRTG 195
Db 455 YHMMWDSRLFNYSWEVLRYLLSNARWLDEYKDFGPRFDGVTSMN--YTHHGLQVSPFG 512
Qy 196 RSIVY 200
Db 513 NYSEY 517
RESULT 8
US-11-057-058-62
Sequence 62, Application US/11057058
Publication No. US2005024400A1
GENERAL INFORMATION:
APPLICANT: LeBowitz, Jonathan
APPLICANT: Maga, John
TITLE OF INVENTION: ACID ALPHA-GLUCOSIDASE AND FRAGMENTS THEREOF
FILE REFERENCE: SYN-011
CURRENT APPLICATION NUMBER: US/11/057,058
CURRENT FILING DATE: 2005-02-10
PRIOR APPLICATION NUMBER: US 60/543,812
PRIOR FILING DATE: 2004-02-10
NUMBER OF SEQ ID NOS: 68
SOFTWARE: PatentIn version 3.3

; SEQ ID NO 62
 ; LENGTH: 1827
 ; TYPE: PRN
 ; ORGANISM: Rabbit sp.
 US-11-057-058-62

Query Match 3.4%; Score 79.5; DB 7; Length 1827;
 Best Local Similarity 18.3%; Pred. No. 15;
 Matches 42; Conservative 29; Mismatches 61; Indels 97; Gaps 10;

QY 143 AGPSPGFGYDIDAQTFADWGV-----DILLKDGCCYCD----- 175
 DB 1360 AAFP-----DPPRNTAEWTRTEILDFFNNYMKFDGLWIDMNEPSSFFVNGTTTNCVNT 1413
 QY 176 -----SLENLADGYKXMSLALN-----RTGRSIVYSCWPLMMPFQKPNYTEIR- 220
 DB 1414 ELYNPPYFPELYKRTDGLHFRTWCMETEHLSDGSSVLHYDVHNLGHSQAKPTDYDALQK 1473
 QY 221 -----QVCNHW-----RNPADIDDSWKSISKILDTWTSFNQERIVDVAGP- 259
 DB 1474 TTCKRGIVISRSTYTAGRWAGHWLGDVYARNWDMKSIIGWMEPSLFG-----ISYTGAD 1529
 QY 260 --CGWNPDMVLV-----TGNP-----GLSNWQVQTM 284
 DB 1530 ICGFFNDSEYHLCTRWLTQLGAFYFPARNHNIQTRRQDPVSNQTFVEM 1578

RESULT 9
 US-10-467-657-8554
 ; Sequence 8554, Application US/10467657
 ; Publication No. US20050260581A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CHIRON SPA
 ; APPLICANT: FONTANA Maria Rita
 ; APPLICANT: PIZZA Mariagrazia
 ; APPLICANT: MASIGNANI Vega
 ; APPLICANT: MONACI Elisabetta
 ; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
 ; FILE REFERENCE:
 ; CURRENT APPLICATION NUMBER: US/10/467,657
 ; CURRENT FILING DATE: 2003-08-11
 ; PRIOR APPLICATION NUMBER: GB-0103424.8
 ; PRIOR FILING DATE: 2001-02-12
 ; NUMBER OF SEQ ID NOS: 9218
 ; SOFTWARE: SeqWin99, version 1.04
 ; SEQ ID NO 8554
 ; LENGTH: 322
 ; TYPE: PRN
 ; ORGANISM: Neisseria gonorrhoeae
 US-10-467-657-8554

Query Match 3.4%; Score 79; DB 6; Length 322;
 Best Local Similarity 20.7%; Pred. No. 1.6;
 Matches 78; Conservative 38; Mismatches 101; Indels 160; Gaps 23;

QY 2 QLRNPELHGLCALRFLALVSDIPGARALDNGLARTPTMGLHWRFMCMNLDQCBERPD 61
 DB 27 QCSDELHWCERAG-----FAGIARLHPFKVKKVHMKWRQRK----- 64
 QY 62 SCISEKLFMEAMELMVSEGKQDAG-----YELLICDCWMAQORDSEGRLOA--- 108
 DB 65 -----HLFQAEWTWREGRUKQALLQEAQDFVL-----DSGLIKSACF 102
 QY 109 -----DQRFPHGIRQLA---NYVHSGKLKL-----GIYADVGNKTCAGFPQSP 149
 DB 103 AKWAKSPIYGLDKHSAREGVAALADKKYVVPKGRNAVWRNRLDLFAQV-----F 151
 QY 150 GYDIDIAQTFADWGVLLKFDGCGYCDSENLADGYKXMSLALNRTGRSIVYSCWPLMVM 209
 DB 152 GYAMPETQVF---GLSVPE-----ASRLKXNLQAPY---YAAHATSRD---SKLPMENW 197
 QY 210 P--FQKPNYTEIRQYCN---HWRNFADIDDSWKSISKILDTWTSFNQERIVDVAGPFGWND 264

; SEQ ID NO 62
 ; LENGTH: 1827
 ; TYPE: PRN
 ; ORGANISM: Rabbit sp.
 US-11-057-058-62

Query Match 3.4%; Score 79.5; DB 7; Length 1827;
 Best Local Similarity 18.3%; Pred. No. 15;
 Matches 42; Conservative 29; Mismatches 61; Indels 97; Gaps 10;

QY 143 AGPSPGFGYDIDAQTFADWGV-----DILLKDGCCYCD----- 175
 DB 1360 AAFP-----DPPRNTAEWTRTEILDFFNNYMKFDGLWIDMNEPSSFFVNGTTTNCVNT 1413
 QY 176 -----SLENLADGYKXMSLALN-----RTGRSIVYSCWPLMMPFQKPNYTEIR- 220
 DB 1414 ELYNPPYFPELYKRTDGLHFRTWCMETEHLSDGSSVLHYDVHNLGHSQAKPTDYDALQK 1473
 QY 221 -----QVCNHW-----RNPADIDDSWKSISKILDTWTSFNQERIVDVAGP- 259
 DB 1474 TTCKRGIVISRSTYTAGRWAGHWLGDVYARNWDMKSIIGWMEPSLFG-----ISYTGAD 1529
 QY 260 --CGWNPDMVLV-----TGNP-----GLSNWQVQTM 284
 DB 1530 ICGFFNDSEYHLCTRWLTQLGAFYFPARNHNIQTRRQDPVSNQTFVEM 1578

RESULT 9
 US-10-467-657-8554
 ; Sequence 8554, Application US/10467657
 ; Publication No. US20050260581A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CHIRON SPA
 ; APPLICANT: FONTANA Maria Rita
 ; APPLICANT: PIZZA Mariagrazia
 ; APPLICANT: MASIGNANI Vega
 ; APPLICANT: MONACI Elisabetta
 ; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
 ; FILE REFERENCE:
 ; CURRENT APPLICATION NUMBER: US/10/467,657
 ; CURRENT FILING DATE: 2003-08-11
 ; PRIOR APPLICATION NUMBER: GB-0103424.8
 ; PRIOR FILING DATE: 2001-02-12
 ; NUMBER OF SEQ ID NOS: 9218
 ; SOFTWARE: SeqWin99, version 1.04
 ; SEQ ID NO 8554
 ; LENGTH: 322
 ; TYPE: PRN
 ; ORGANISM: Neisseria gonorrhoeae
 US-10-467-657-8554

Query Match 3.4%; Score 79; DB 6; Length 322;
 Best Local Similarity 20.7%; Pred. No. 1.6;
 Matches 78; Conservative 38; Mismatches 101; Indels 160; Gaps 23;

QY 2 QLRNPELHGLCALRFLALVSDIPGARALDNGLARTPTMGLHWRFMCMNLDQCBERPD 61
 DB 27 QCSDELHWCERAG-----FAGIARLHPFKVKKVHMKWRQRK----- 64
 QY 62 SCISEKLFMEAMELMVSEGKQDAG-----YELLICDCWMAQORDSEGRLOA--- 108
 DB 65 -----HLFQAEWTWREGRUKQALLQEAQDFVL-----DSGLIKSACF 102
 QY 109 -----DQRFPHGIRQLA---NYVHSGKLKL-----GIYADVGNKTCAGFPQSP 149
 DB 103 AKWAKSPIYGLDKHSAREGVAALADKKYVVPKGRNAVWRNRLDLFAQV-----F 151
 QY 150 GYDIDIAQTFADWGVLLKFDGCGYCDSENLADGYKXMSLALNRTGRSIVYSCWPLMVM 209
 DB 152 GYAMPETQVF---GLSVPE-----ASRLKXNLQAPY---YAAHATSRD---SKLPMENW 197
 QY 210 P--FQKPNYTEIRQYCN---HWRNFADIDDSWKSISKILDTWTSFNQERIVDVAGPFGWND 264

```

; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-236

Query Match
Best Local Similarity 3.4%; Score 78.5; DB 6; Length 622;
Matches 65; Conservative 24; Mismatches 94; Indels 121; Gaps 16;

QY 97 APQRSEGRLOADPQRFPHGIRQLANYVHSGKLGIYADVGNKTCAGFPGSFG-----150
Db ADYRNAQGRILSGPE-----YLTV-----FKGDTGAELTTV 284
QY 151 YYDIDAQTFADWG-----VDLLKPDGCGYCDLEN---LADGY---KMSLALNRTGRSIV 199
Db NYEPARGNADWDGSGYGRVDRFLAGVAYLDGERPSFVMARGYYTRTVLVAYNFRGGKIT 344
QY 200 YSCEWPLYWPPQK---PNVTEIRQCNHWRNFADIDDSWKSISILDWTSFNOERI---253
Db 345 -----KLWTFDSAPGNGAYAGQGNHSLSVADVDDGDK-----DELIYCAM 385
QY 254 -VDVAGPG-----GWNDDPMLVIGN-----FGLSWNQOVTQMALWA 288
Db 386 AVDHQKGLYSTGWGHDAMHTGNDLPSRPGLEVQVHENSNSPYGLSFRDAKTGKIWG 445
QY 289 I-----MAAPLFMSNDRHISPOAKA---LLQDKDVTAINDPDKQKQYQLRQGDN 336
Db 446 VHAGKDVGRGMAADI-----DPYEGAENVANGSLYTAKGV-----KIGNTLPSSTN 492
QY 337 FEVW 340
Db 493 FGIW 496

RESULT 12
US-10-821-234-1141
; Sequence 1141, Application US/10821234
; Publication No. US200502551141A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt seq_genes Version 1.0
; SEQ ID NO 1141
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1141

Query Match
Best Local Similarity 3.4%; Score 78; DB 6; Length 760;
Matches 40; Conservative 25; Mismatches 59; Indels 84; Gaps 8;

QY 49 RPMCNDLQCEPDSCISEKLFMEMAELMVSEGWDAGYEYLCIDDCWM-----96
Db 362 RIVTVYFSAHDNDANLQMDIFLEEDGNVNSKLTGDSVMYHWCNEAWTRPDLVPVFGW 421
QY 97 -----APQSDSGRLQADP---ORFPHGIRQLANYVHSGKLGIYADVGNKTCAGFPGS 148
Db 422 QAVDSTPQNSDGMVRCGPASVQATKHG-----HVCFQDAP 458
QY 149 FGYYDIDAQTFADWGVDLL---KFDGCVY-----DSLENLADGY 184
Db 459 F-----VFAEVNSDLIYITAKDKGTHVENVATHIGKLIYVTKQIGGDMIDITDY 510
QY 185 KHMS-----LALNRTGRSIVSCWPL 206

; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-236

Query Match
Best Local Similarity 3.4%; Score 78; DB 6; Length 786;
Matches 76; Conservative 39; Mismatches 106; Indels 180; Gaps 20;

QY 5 NPETHLGCALALRFLALVSDIPGARLDNGLAR-----38
Db 280 NGYLHIGHAKAM-----FVDFGLAKERGCGCYLRVDDTNPEAEKEEYINHI 325
QY 39 TPTWGLHWRFCNLDQEE-----PDSCISE--K 67
Db 326 EEIVKMGWEPPKELYDLAVELIRGHAYVDHQTADRIKEYREKQNSFWDRPIEESLK 385
QY 68 LFMEMAELMVSEGWDAGYEYLCIDDCWMAPQDSEGRLOADPQRFPHGIRQLANYVHSG 127
Db 386 LFDWMRGIIIEG-----KATLRMKQDMQSDNFMYDLIAY----421
QY 128 GLKLGIVADVGNKTCAGFPGSFGYVDIDAQTFADWGVDLLKPDGCGYCDLENLADGYKHM 187
Db 422 RIKFAPHPKAGDKWCI-YPS---YD-----YAHCTVDSLENIT---HS 457
QY 188 SLALN-RTGRSIVSCWPL-----YMWPFQKPNVTEIRQYCNHWRNFADIDDSWKS 238
Db 458 LCTLEFETRASY--WLLHSLSLYMPYVWYWEYGRNLVNT-----493
QY 239 IKSILDWTSFN---QERIVDVAGPGWNDPMLVIGNFGLSWNQOVTQMALWAIM-----290
Db 494 -NTVNSKRLNYIYTNKYVD-----GWDDPRLTILS--GLR-RRGVTSTAINAFVRGIGI 544
QY 291 ----AAPLFMSNDRHI-----SPOKALLQDKDVIAIN 320
Db 545 TRSDGSMIHVSRLHREELNKTAPRTMVLNPLKVVITN 585

RESULT 14
US-10-131-826A-450
; Sequence 450, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 26, 2005, 13:48:48 ; Search time 282 Seconds
(without alignments)
774.752 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MOLRRPELHLCALALRFLA.....RLRSHNPTGTVLLQLENTM 421

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4168288 seqs, 259477437 residues

Total number of hits satisfying chosen parameters: 8336576

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO pool_p/US10602219/runat_23122005_151144_17285/app_query.fasta_1.583
-DB=Published Applications NA New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-MAXLEN=2000000000 -USER=US10602219 @CGN 1.1 183 @runat_23122005_151144_17285
-NCPUS=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA New.*
1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US12_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US13_NEW_PUB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US14_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US15_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	99	4.3	1515	7 US-11-073-112-5	Sequence 5, Appli
2	95.5	4.1	268681	6 US-10-933-025-22	Sequence 22, Appl
3	94	4.1	1241	6 US-10-750-185-56907	Sequence 56907, A
4	93	4.0	1082144	7 US-11-117-187-211	Sequence 211, App
5	91.5	4.0	2177	7 US-11-073-112-4	Sequence 4, Appli
6	90.5	3.9	2716	6 US-10-750-185-31179	Sequence 31179, A
7	89.5	3.9	1766	6 US-10-750-185-35182	Sequence 35182, A
8	88	3.8	98345	7 US-11-112-908-36	Sequence 36, Appli

9	88	3.8	127340	7 US-11-112-908-35	Sequence 35, Appl
10	87.5	3.8	2576	9 US-11-038-284-16	Sequence 16, Appl
11	87.5	3.8	2578	9 US-11-038-284-19	Sequence 19, Appl
12	87.5	3.8	5301	7 US-11-108-528-83	Sequence 83, Appl
13	86.5	3.7	1615	6 US-10-750-185-56545	Sequence 56545, A
14	85	3.7	1212	6 US-10-750-185-55673	Sequence 55673, A
15	85	3.7	1314	9 US-11-073-626-4	Sequence 4, Appli
16	84.5	3.7	3033	9 US-11-038-284-14	Sequence 14, Appl
17	84.5	3.7	3033	9 US-11-038-284-41	Sequence 41, Appl
18	84	3.6	2866	6 US-10-510-386-235	Sequence 235, App
19	83.5	3.6	948	7 US-11-055-822-1	Sequence 1, Appli
20	83.5	3.6	1262	6 US-10-750-185-45064	Sequence 45064, A
21	82	3.5	1314	9 US-11-073-626-2	Sequence 2, Appli
22	82	3.5	2626	6 US-10-750-185-58314	Sequence 58314, A
23	81.5	3.5	889	6 US-10-750-185-37410	Sequence 37410, A
24	81.5	3.5	1335	6 US-10-467-657-8553	Sequence 8553, Ap
25	81.5	3.5	1335	6 US-10-374-954-4	Sequence 4, Appli
26	81.5	3.5	1548	6 US-10-630-203-5	Sequence 5, Appli
27	81.5	3.5	1548	6 US-11-103-037-6	Sequence 6, Appli
28	81.5	3.5	2380	7 US-11-103-240-161	Sequence 161, App
29	81.5	3.5	2399	6 US-10-750-185-35565	Sequence 35565, A
30	81.5	3.5	2498	6 US-10-750-185-45595	Sequence 45595, A
31	81.5	3.5	189252	7 US-11-121-086-54	Sequence 54, Appl
32	81	3.5	930	7 US-11-074-176-107	Sequence 107, Appl
33	81	3.5	4860	6 US-10-971-982-1	Sequence 1, Appli
34	81	3.5	168516	7 US-11-121-086-3	Sequence 3, Appli
35	80.5	3.5	1242	6 US-10-467-657-31	Sequence 31, Appl
36	80.5	3.5	1242	6 US-10-467-657-8371	Sequence 8371, Ap
37	80.5	3.5	6423	6 US-10-750-185-61474	Sequence 61474, A
38	80	3.5	1559	6 US-10-864-758-1	Sequence 1, Appli
39	80	3.5	1611	6 US-10-864-758-6	Sequence 6, Appli
40	80	3.5	5745	6 US-10-750-185-54380	Sequence 54380, A
41	80	3.5	35344	6 US-10-995-561-13307	Sequence 13307, A
42	80	3.5	60844	6 US-10-995-561-13359	Sequence 13359, A
43	79.5	3.4	867	6 US-10-454-437-363	Sequence 363, App
44	79.5	3.4	960	6 US-10-467-657-3253	Sequence 3253, Ap
45	79.5	3.4	1251	9 US-11-082-389-121	Sequence 121, App

ALIGNMENTS

RESULT 1
US-11-073-112-5
; Sequence 5, Application US/11073112
; Publication No. US20050260627A1
; GENERAL INFORMATION:
; APPLICANT: Hintz et al.
; TITLE OF INVENTION: Mannosidases and Methods for using the Same
; FILE REFERENCE: 62447-02
; CURRENT APPLICATION NUMBER: US/11/073,112
; CURRENT FILING DATE: 2005-03-03
; PRIOR APPLICATION NUMBER: US 10/089,211
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: PCT/US00/27210
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/157,341
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5
; LENGTH: 1515
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-11-073-112-5

Alignment Scores:
Pred. No.: 0.492 Length: 1515
Score: 99.00 Matches: 62
Percent Similarity: 35.74% Conservative: 32
Best Local Similarity: 23.57% Mismatches: 74
Query Match: 4.28% Indels: 95
DB: 7 Gaps: 18


```
Db 985 ---GGCTTCTACTGCTGCTGGTCTCCGGGAGCAGCAGCTGCAGCGCTGGACC 1041
      |||::: |||
      :::
Qy 341 uArgProLeuSerGlyLeuAlaTrpAlaValalaMet----- 353
      |||::: |||
      :::
Db 1042 GGGTCCCTGTCAAGGCTCGTGTGGGAGCAGCTGCTTGACAGCAGCTTCCTGTT 1101
      |||::: |||
      :::
Qy 354 -----lleAsnArg-----GlnGluIleGlyG1 361
      :::: |||
      :::
Db 1102 CACTGGGGGGTGGTGGAGCGCTGGGAGCAGGTGACAGGTGGAGCGGCAATGGGT 1161
      :::: |||
      :::
Qy 361 yProArgSerTyrThrIleAlaValala-----SerLeuGlyLysGlyVa 376
      |||::: |||
      :::
Db 1162 TCCTAGGCTCAAAACATCACCCAGGCTGGAGGAGCAGCAATAAGCTTCAACTTCCCGT 1221
      |||::: |||
      :::
Qy 376 lalaCysAsnProAlaCys 382
      :::: |||
      :::
Db 1222 GCCTCTCTCCCTAGGTGT 1240
      :::: |||
      :::

RESULT 4
US-11-117-187-211
; Sequence 211, Application US/11117187
; Publication No. US20050266560A1
; GENERAL INFORMATION:
; APPLICANT: PREUSS, DAPHNE
; APPLICANT: COPENHAVEN, GREGORY
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
; FILE REFERENCE: ARCD:309US
; CURRENT APPLICATION NUMBER: US/11/117,187
; PRIOR FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: US/09/531,120
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,219
; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 211
; LENGTH: 1082144
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-11-117-187-211

Alignment Scores:
Pred. No.: 3 94e+04 Length: 1082144
Score: 93.00 Matches: 76
Percent Similarity: 30.97% Conservative: 42
Best Local Similarity: 19.95% Mismatches: 117
Query Match: 4.02% Indels: 146
DB: 7 Gaps: 20

US-10-602-219-12 (1-421) x US-11-117-187-211 (1-1082144)
Qy 3 LeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAlaLeuVal 22
      |||::: |||
      :::
Db 565984 CTGCGTTATCCGAGGTAGAA--GGGGCGCGCGCGCTTAGAATAGTCGGA----- 566034
      |||::: |||
      :::
Qy 23 SerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMet 42
      |||::: |||
      :::
Db 566035 -----GATCTCCGAAACCTCGTTTGTGAT-----TTCACACCTCGCATG 566076
      :::: |||
      :::
Qy 43 GlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluProAspSer 62
      :::: |||
      :::
Db 566077 GACTTGATGACTGGCTTGTCTTCCTTTGGTTTCAATCGATAACGTTCCGATCAG 566136
      :::: |||
      :::
Qy 63 CysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLys 82
      |||::: |||
      :::
Db 566137 -----CGAGAGAATCTCGTTCTTCACTTAGCT----- 566163
      :::: |||
      :::
Qy 83 AspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSer 102
      :::: |||
      :::
Db 566164 -----AACTCT 566169
      :::: |||
      :::
Qy 103 GluGlyArgLeuGlnAlaAspProGlnArgPheProHisGly----- 116
```

RESULT 5

```
US-11-073-112-4
; Sequence 4, Application US/11073112
; Publication No. US20050260627A1
; GENERAL INFORMATION:
; APPLICANT: Hintz et al.
; TITLE OF INVENTION: Mannosidases and Methods for using the Same
; FILE REFERENCE: 62447-02
; CURRENT APPLICATION NUMBER: US/11/073,112
; CURRENT FILING DATE: 2005-03-03
; PRIOR APPLICATION NUMBER: US 10/089,211
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: PCT/US00/27210
; PRIOR FILING DATE: 2000-10-02
```

; PRIOR APPLICATION NUMBER: 60/157,341
 ; PRIOR FILING DATE: 1999-10-01
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: PatentIn version 3.3
 ; SEQ ID NO 4
 ; LENGTH: 2177
 ; TYPE: DNA
 ; ORGANISM: Aspergillus nidulans
 US-11-073-112-4

Alignment Scores:
 Pred. No.: 6.48 Length: 2177
 Score: 91.50 Matches: 45
 Percent Similarity: 41.72% Conservative: 23
 Best Local Similarity: 27.61% Mismatches: 49
 Query Match: 3.95% Indels: 46
 DB: 7 Gaps: 11

US-10-602-219-12 (1-421) x US-11-073-112-4 (1-2177)

QY 27 ProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrpLeuHis 46
 DB 1059 CCGGAGCCAGCGCAGTA-----CGAACGTTCCCTGGATTGAT 1097
 QY 47 TrpGluArgPheMet---CysAsnLeuAspCysGlnGluGluProAspSerCysIleSer 65
 DB 1098 TGGAGCGCAGTCAATATTGCCGACGCGCAAGCTCGCAA-----TGGTCA 1142
 QY 66 GluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLys-AspAlaG1 85
 DB 1143 CATCAGCTG-----GAATGGTGGCGAGCTCGTA 1172
 QY 85 YTrpGluTrpLeu-----CysIleAspAs 93
 DB 1173 CTACGAGTACCTGATCAAGATGACTAGCTACGATCCGACGCTTGGCTCTACCGGA 1232
 QY 93 pCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGlnArgPh 113
 DB 1233 CGCTGGGTGCGAGCTGCGAGTCGAGTCAGCATCAACCATCTGGCTTCGACCCGTCACCCG 1292
 QY 113 eProHisGlyIleArgGlnLeuAlaAsnTyValHisSerLysGlyLeuLysLeuGlyI1 133
 DB 1293 CCCA---GACGTGACTTCTTGGCCACTTAC-----AACGAGGAGCATCAGCTGGGCT 1343
 QY 133 eTyAlaAspValGlyAsnLysThrCysAlaGlyPheProGlySerPhe-----GlyTy 151
 DB 1344 GACACGCCAA-----CACCTGACCTGCTTCGAC---GGTGAAGCTTCTGCTGGTGG 1394
 QY 151 rTyAspIleAspAlaGlnThrPheAlaAspTrpGlyValAspLeuLysPheAspG1 171
 DB 1395 GACATTGTGGACCGCAGGACTTTGTGCACTTCGGCTTGACCTTGTCGC-----GG 1448
 QY 171 yCysTy 173
 DB 1449 CTGCCAC 1455

RESULT 6

US-10-750-185-31179
 ; Sequence 31179, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.
 ; APPLICANT: Denise, Sue K.
 ; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom
 ; APPLICANT: FANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 ; FILE REFERENCE: MM1100-2
 ; CURRENT APPLICATION NUMBER: US/10750,185
 ; CURRENT FILING DATE: 2003-12-31
 ; PRIOR APPLICATION NUMBER: US 60/437,482

; PRIOR FILING DATE: 2002-12-31
 ; NUMBER OF SEQ ID NOS: 64922
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 31179
 ; LENGTH: 2716
 ; TYPE: DNA
 ; ORGANISM: Bovine
 US-10-750-185-31179

Alignment Scores:
 Pred. No.: 11.8 Length: 2716
 Score: 90.50 Matches: 48
 Percent Similarity: 30.84% Conservative: 22
 Best Local Similarity: 21.15% Mismatches: 82
 Query Match: 3.91% Indels: 76
 DB: 6 Gaps: 11

US-10-602-219-12 (1-421) x US-10-750-185-31179 (1-2716)

QY 38 ArgThrProThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCys--- 56
 DB 1674 AGGTTCCAGCCTCGGACCCCTCAGCTGGTAAAGATC-----CACCTGCAATGCGGG 1727
 QY 57 GlnGluGluProAspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMet 76
 DB 1728 AGACTGAGTTCGATCCCTGCTGGGAAG----- 1757
 QY 77 ValSerGluGlyTrpLysAspAlaGlyTyTrpGluTyLeuCysIleAspAspCysTrpMet 96
 DB 1758 -----ATCCCTGGAGAGGGAGGCTCCCACTCTGGCTGGAGATTCATGGACT 1811
 QY 97 AlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGly 116
 DB 1812 GTATAGTCCATAGGTGCGAGTCGAGTCGACAA-----CCAAGCGACTTT 1856
 QY 117 IleArgGlnLeuAlaAsnTyValHisSerLysGlyLeuLysLeuGlyIleTyAlaAsp 136
 DB 1857 -----CACTTCATTCGGGACAAAGGGCTGGTATTGGCGGTT 1892
 QY 137 ValGlyAsnLysThrCysAlaGlyPheProGlySerPheGlyTyTrpAspIleAspAla 156
 DB 1893 -----TGTACAGTTCCTCTCTGTCTTGGCTGGGCAAGCGGATGCT 1937
 QY 157 GlnThrPheAlaAspTrpGlyValAspLeuLysPheAspGlyCysTyTrpCysAspSer 176
 DB 1938 ACCTCTGTGGCTC-TGG-----TGCACTGT 1963
 QY 177 LeuGluAsnLeuAlaAspGlyTyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArg 196
 DB 1964 GTGCAAGGTGCTGGCTGATGGA-----GGTGTCTAGGTGGGAGGTGAGGACTG 2011
 QY 197 SerIleValTyTrpSerCys-----GluTrpProLeu 206
 DB 2012 GGAGTGGCCAGGGCTGTGGCAGGGCCCTGACAGCCTCTGCCACTCTCCTGACACAG 2071
 QY 207 TyrMetTrpProPheGlnLys-----ProAsnTy 216
 DB 2072 GCATGGTGGCCAGCCAGCTCCCTAGGAGTGTGCTCTCTGCCAAGCCCTCTTGTCCCGCTC 2131
 QY 217 ThrGluIleArgGlnTyTr-----CysAsnHisTrpArgAsnPhe 229
 DB 2132 ACGTCTCTCCAGGCCACGAAGCTGCCACCACCTCTCTCTGCTGCCCTGGGAGACCTT 2191
 QY 230 AlaAspIleAspAspSerTrp 236
 DB 2192 TCCCACTGGGACCACTCATGG 2212

RESULT 7

US-10-750-185-35182
 ; Sequence 35182, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.

APPLICANT: DeNISE, Sue K.
APPLICANT: KERR, Richard
APPLICANT: ROSENFELD, David
APPLICANT: HOLM, Tom
APPLICANT: BATES, Stephen
APPLICANT: FANTIN, Dennis
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
FILE REFERENCE: MM1100-2
CURRENT APPLICATION NUMBER: US/10/750,185
CURRENT FILING DATE: 2003-12-31
PRIOR APPLICATION NUMBER: US 60/437,482
PRIOR FILING DATE: 2002-12-31
NUMBER OF SEQ ID NOS: 64922
SOFTWARE: PatentIn version 3.1
SEQ ID NO 35182
LENGTH: 1766
TYPE: DNA
ORGANISM: Bovine 19866880928175
US-10-750-185-35182

Alignment Scores:
Pred. No.: 8.23 Length: 1766
Score: 89.50 Matches: 57
Percent Similarity: 31.33% Conservative: 21
Best Local Similarity: 22.89% Mismatches: 85
Query Match: 3.87% Indels: 86
DB: 6 Gaps: 15

US-10-602-219-12 (1-421) x US-10-750-185-35182 (1-1766)

QY 21 LeuValSerTIPAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAla----- 37
DB 481 CTCTGTCATGGATTTCCAGGCAAG-----AATCTGGATCATGTCGCAATTC 531
QY 38 -----ArgThrProThrMetGly-----TrpLeuHisTrpGlu-Ar 49
DB 532 CTCTCCAGGAGATCTTGCCAAACCCCAAGGATCGAACCCATTTCTCTCACTGGGAGGCG 591
QY 49 gPheMetCysAsnLeuAspCysGlnGluProAspSerCysIleSerGlyLeuPh 69
DB 592 GATTCCTTCCACTAGCCAGCCAGGGAAGCCCCAGTAGCTGTGTAGCACATTTTATA 651
QY 69 eMetGluMetAlaGluLeuMetValSerGluGlyTrpLysAsp-----AlaG1 85
DB 652 CTTTCAA-----CAGAAAGGTGGGAGAAACAGTCTTGGGGGAGG 693
QY 85 YTrpGluTrpLeu---CysIleAspAsp-----CysTrpMetAlaProGlnArgAspSe 102
DB 694 ATATGCCCTCTGCTCTGTGTCTTCTCTCTGTTCTCTGAGGCTGGACATCAGG----- 748
QY 102 rGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeuAlaAs 122
DB 749 -GAAGGGATGCTGGTGCCTCT----- 769
QY 122 nTyrrValHisSerLysGlyLeuLysLeuGlyIleTyrrAlaAspValGlyAsnLysThrCy 142
DB 770 -----GATGTCAGAGCCAGCTTGGCCACTTAAGTCAGGAGGAACTGCAGACTG 822
QY 142 sAlaGlyPheProGlySerPheGlyTyrrTyrrAspIleAspAlaGlnThrPheAlaAspTr 162
DB 823 CCAGGGGCTGACAGATCATCGGA-----GGCTG 852
QY 162 pGlyValAspLeuLeuLys-----PheAspGlyCysTyrrCysAs 175
DB 853 GAGAACTGGGCTTACAAAGAACAGGAGCAATCTTGAGGCTGGGATGCAATGGGCAGA 912
QY 175 pSerLeu-----GluAsnLeuAlaAspG1 183
DB 913 TGCCTCTAGGGGAAGAGGCTGGGTAGTGGGTGACTTCTTAGGAAGAGGCCACTGG 972
QY 183 YTrpLysHisMetSerLeuAlaLeuAsnArgThr-----GlyArgSerIleVa 199
DB 973 CTATGTCAGTGTGCCATCATCCCGACCCAGGATTTTAGTTTCCAGGGGGAGAGATCCACT 1032

QY 199 lTyrrSerCys---GluTrpProLeuTyrrMetTrpProPheGlnLysProAsnTyrrThrG1 218
DB 1033 TGCCTTCTGTGTGCTCTGGGTGTC-----CCTTCGCTCTACCA 1071
QY 218 uIleArgGlnTyrrCysAsnHisTrp 226
DB 1072 GACTGAACCTGTAGGAGGACACTGG 1096
RESULT 8
US-11-112-908-36
Sequence 36, Application US/11112908
Publication No. US20050260659A1
GENERAL INFORMATION:
APPLICANT: Harris, Cole
APPLICANT: Davis, Lisa M.
TITLE OF INVENTION: Breast Cancer Biomarkers
FILE REFERENCE: 04-164-US
CURRENT APPLICATION NUMBER: US/11/112,908
CURRENT FILING DATE: 2005-04-22
PRIOR APPLICATION NUMBER: US 60/564,758
PRIOR FILING DATE: 2004-04-23
PRIOR APPLICATION NUMBER: US 60/575,978
PRIOR FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/631,702
PRIOR FILING DATE: 2004-11-30
PRIOR APPLICATION NUMBER: US 60/633,826
PRIOR FILING DATE: 2004-12-07
NUMBER OF SEQ ID NOS: 511
SOFTWARE: PatentIn version 3.3
SEQ ID NO 36
LENGTH: 98345
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (49350)..(49350)
OTHER INFORMATION: n is a, c, g, or t
US-11-112-908-36
Alignment Scores:
Pred. No.: 4.55e+03 Length: 98345
Score: 88.00 Matches: 49
Percent Similarity: 32.4% Conservative: 15
Best Local Similarity: 24.8% Mismatches: 75
Query Match: 3.80% Indels: 58
DB: 7 Gaps: 11
US-10-602-219-12 (1-421) x US-11-112-908-36 (1-98345)
QY 225 HisTrpArgAsnPheAlaAspIleAspSerTrpLysSerIleLysSerIleLeuAsp 244
DB 76964 CATTCGCGA----- 76972
QY 245 TrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGlyGlyTrpAsnAsp 264
DB 76973 -----CAGAGCGGGCAGCGCCGCCCTCGGGCCGAGAGG----- 77008
QY 265 ProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsn-----GlnGlnValThr 282
DB 77009 ---GAGCGGTACGAGCGGGGGCGCTGGCACCGCGGTGGAACTCGCGCGGGGTCC 77065
QY 283 GlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHis 302
DB 77066 GCCCTGGCCCTCTGCCCGCCGCCCGCTGCTTTGCTTCTCG----- 77110
QY 303 IleSerProGlnAlaLysAlaLeuLeuGlnAspValIleAlaIleAsnGlnAsp 322
DB 77111 CTCGCCCTCTCCCGCCCGCTCGCTTCACAGCGCGCGGAGGACCTAACGCCGGT 77170
QY 323 ProLeuGlyLysGlnGlyTyrrGlnLeuArgGlnGlyAspAsnPheLeuValTrpGluArg 342
DB 77171 CCTCTAGGAACCTCGGGCGGGCAGCACCCCGGGGA-----TTCTGCTGGGTCTCT 77221


```
Db 2051 ||||| :||| ||| ||| :|||: 2104
Qy 64 CACTGGTCAAGAGCTTCTGCTTTGGATTGATGGACAGCAACCGAAT
Db 2105 -----TTATTGGACTGATATACCTACCTACAGACCATCAGCAG 2140
Qy 84 AlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSerGlu 103
Db 2141 AGCCTTTATGAATGAGTGCAGTGCACCTGCAACATGT----- 2173
Qy 104 GlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleAetGlnLeuAlaAsnTyr 123
Db 2174 GGTAGAAATTCGGCTTAGAATTATCCAGA---AGGCATGGCAGTAGA----- 2215
Qy 124 ValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsn-----Lys 140
Db 2216 CTGGCTTGGGAGAACTTGCTACTGGCGACACACAGGACGATCGAATGAGTGTCTCAA 2275
Qy 141 ThrCysAlaGlyPheProGlySerPheGly-TyrTyrAspIleAspAla----- 156
Db 2276 GTTGATGGGACGACCGACAAAGTTTGTGTGGAAAGACCTAGATAGTCCAGAGCTCT 2335
Qy 157 -----GlnThrPheAlaAspTrpGly-----Va 164
Db 2336 CGCGTTGACCTGCGGAGGATTTATGATTGGACTGAATGGGTGGAACCTTAAGAT 2395
Qy 164 LaspLeuLeuLysPheAspGlyCysTyrCysAspSerLeu----- 177
Db 2396 AGACAGAGCTGCAATGGATGGAGTGAAGTACTACTTACCTTAGTTCCAATGTGGGCGGCG 2455
Qy 178 -GluAsnLeuAlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSe 197
Db 2456 AAACGGCTTAATGATTGATTATGCTAAAGAGGCTTTA-TTGGACAGACCTGGACACCA 2514
Qy 197 rle-----ValTyrSerCysGluTrpPro----- 205
Db 2515 ACTTAATAGAAATCTTCAAAATATGCTTGGGCTCAACCGTGAAGTTATAGCAGATGACTTGC 2574
Qy 206 -----LeuTyrMetTrpTrpProPheGlnLysProAs 215
Db 2575 CTCATCCTTTTGGCTTAAGTACCTAGTACCAAGATTATATCTACTGGAGGAGTGGACCCGAC 2634
Qy 215 nTyrThrGluLeuArgGlnTyrCysAsnHisTrp-ArgAsnPheAlaAspIleAspAsps 235
Db 2635 CGACGATTGAGCGTCCCAACAA--AACCAGTGGCCAAACCGCACCATCATTCAGGGCC 2691
Qy 235 erTrpLysSerLysSerLysLeuAspTrpTrpSerPheAsnGlnGluArgIleVala 255
Db 2692 ATTTGGATTATGATGGACATCTCTGCTTTTCACTCATCTCGACAGTCCAGGTTGGAATG 2751
Qy 255 spValAlaGlyProGlyTyrTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyL 275
Db 2752 AATGTCTTCCAGCAATGGGCACTGCTC-CCACCTCTGCTTGGCTGTGCGAGTGGGG- 2809
Qy 275 euserTrpAsnGlnGlnValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuP 295
Db 2810 -----TTTTGTTGTGG-----ATGCCCTGCCCATAC 2837
Qy 295 he-MetSerAsnAspLeuArgHisIleSerProGlnAlaLysAlaLeu----- 311
Db 2838 TCTCTTAATGCTGACAAACAGGACTTGTAGTGTCTCTACGACTTTCTGCTCTTCAGTCAA 2897
Qy 312 -----GlnAspLysAspValIle----- 317
Db 2898 AAGAGTGCCATCAACGGCATGTGTGATGATGAACACAGAGCCCGACATCATCTCTCC 2957
Qy 318 -----AlaIleAsnGlnAspProLeuGlyLysGlnGlyTyr 329
Db 2958 ATCCACAGCCTTCGGAAATGTCGGGCAATTGACTATGACCCCACTGGACAGCAACTCTAT 3017
Qy 330 GlnLeuArgGlnGlyAspAsnPheGluValTrpGluArgProLeuSerGlyLeuAlaTrp 349
```

```
Db 3018 TGGATTGACTCACGACAAAAC----- 3038
Qy 350 AlaValAlaMetIleAsnArgGlnGlnIleGlyProArgSerTyrThrIleAlaVal 369
Db 3039 -----ATGATCCGAAAGGACACAGAGATGGCAGCGGCTTACTGTGGTTGTG 3089
Qy 370 AlaSerLeuGlyLysGlyValAlaCysAsnProAlaCysPheIleThrGlnLeuLeuPro 389
Db 3090 AGCTCAGTT-----CCGAGTCAGAACCTG---GAATACAAACCC 3125
Qy 390 VallysArgLysLeuGlyPheTyrGlu-----TrpThrSerArgLeuArgSer 405
Db 3126 TATGACCTCAGATTGATATTACAGCGCTACATCTACTGGACTTGTGAGCTACCAAT 3185
Qy 406 HisIleAsnProThr-----GlyThrValLeu 414
Db 3186 GTCATTATGTGACAAAGATTAGATGGGAGATCAGTTGGAGTGGTCTG 3233

RESULT 13
US-10-750-185-56545/c
; Sequence 56545, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56545
; LENGTH: 1615
; TYPE: DNA
; ORGANISM: Bovine 19866880768439
US-10-750-185-56545

Alignment Scores:
Pred. No.: 16.4 Length: 1615
Score: 86.50 Matches: 94
Percent Similarity: 21.52% Conservative: 51
Best Local Similarity: 20.43% Mismatches: 173
Query Match: 3.74% Indels: 144
DB: 6 Gaps: 21

US-10-602-219-12 (1-421) x US-10-750-185-56545 (1-1615)
Qy 23 SerTrpAspIleProGlyAlaArgAlaLeuAspAsnGly----- 35
Db 1435 TCCTGTAGTCTCCTGGGCTGATTCACAGGGCAGTGGCTCTGTGGTGGGGTGGATTG 1376
Qy 36 -----LeuAlaArgThrProThrMet 42
Db 1375 CAGGAAATGGAGCAAAAATTTCTCCCAAGTCAGGAGAAAGCAGAGTTCCCAACTCTC 1316
Qy 43 GlyTrpLeuHisTrp---GluArgPheMetCysAsnLeuAspCysGlnGlu----- 59
Db 1315 AGCCACAGAGGTGGCATGAGATTGTCTACTGCCTCTGAAATTGTTGAAGCAGGTTCCG 1256
Qy 60 -----ProAspSerCysIleSerGlnLysLeuPheMetGlu 71
Db 1255 CCATCTCTCTCTCGGCGCACCTGAGCCCAATGTCATGCATCATGAAGGGGCTTTTATG 1196
Qy 72 MetAlaGluLeuMetValSerGlnGlyTrpLysAspAlaGlyTyrGluTyrLeuCysIle 91
Db 1195 AGATTTCATGAGGCCCTAGGGGGAATAATTGGCATTTCTTCT-----TAC 1154
```

```
QY 92 AspAspCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspPro--- 110
Db 1153 GATGATTTATTTATGAGCGCATGGATGATGAGGATAGAGGCTGCTCCATTA 1094
QY 111 GlnArgPheProHisGlyIleArgGlnLeuAlaAsnTrpValHisSerLys----- 127
Db 1093 GTGAGAGCGCACAGGAGGAGGCTGCTGCTGCTTTGA-CATTTCAAGAGCGCATC 1035
QY 127 ----- 127
Db 1034 CTACGAGGAGCTGCTTTCTTTTAGAGGTTTACCTTTAATGTTCTATTATCATGGCCT 975
QY 128 -----GlyLeuLysLeuGlyIleTrp 134
Db 974 TGTCTGAGCAGCTGGAGAGGTAGAGTAGCGGTGCGGGGGCTTTGGGGGGAGACAG 915
QY 135 AlaAspValGlyAsnLysThrCysAla---GlyPheProGlySerPheGlyTrpTrpAsp 153
Db 914 GCGGGGTGGTCACTCAGTGGCCAGCTTGGGGCGCGGAGGCTGGGACCGGGCG 855
QY 154 IleAspAlaGlnThrPheAlaAspTrpGlyValAspLeuLysPheAspGlyCysTrp 173
Db 854 AGAGCTGTGTCACG-----GACTGG----- 834
QY 174 CysAspSerLeuGlnAsnLeuAlaAspGlyTrpLysHisMet---SerLeuAlaLeuAsn 192
Db 833 -----CAGCCCGCGCGCCAGCAGCGCCACCTCGCGGCTTGAGCTGCAGG 786
QY 193 ArgThrGlyArgSerIleValTrpSerCysGluTrpProLeuTrpMetTrpProPheGln 212
Db 785 CGCGCGGGCAGCAGCT-GTTTACTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCG 733
QY 213 LysProAsnTrpTrpGluIleArgGlnTrpCysAsnHisTrpArgAsnPheAlaAspIle 232
Db 732 ---CCGGCTCCACCGCAGCAGCAGCTTTCACATCTCTGG----- 694
QY 233 AspAspSerTrpLysSerIleLysSerIleLeuAspTrpThrSerPheAsnGlnGluArg 252
Db 693 -----CTCTGGCGCTGGCGCAAGCCCTTTTCTCACTATGAGCTCTGGGGCTGTGGACAGA 640
QY 253 IleValAspValAlaGlyProGlyGlyTrp-----AsnAspProAspMetLe 268
Db 639 ATCCCAAACTTGAG---CCCGGGCTGGATTGCTTCGGTCTCTGCTGACTGACTCAT 583
QY 268 uValIleGlyAsnPheGlyLeuSerTrpAsnGln---GlnValThrGlnMetAlaLeuTr 287
Db 582 AGGACCCCGAGCGCTGGGCTCAGTTTCCCATGTGTCCATATGTCACATAATCAGATTGATTTG 523
QY 287 palalie-MetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSerProGlnA 307
Db 522 GCAAAATGTTTGGCAGTAGAGAGATTCTTCAAAATGATGTCCAGTGTACAGGCCCAAGAAT 463
QY 307 lalysAlaLeuLeuGlnAspLysAspValIleAlaIleAsnGlnAspProLeuGlyLysG 327
Db 462 ACAAGCA-----GATAGCTCAGGTGGCC 439
QY 327 lngGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrpGluArgProLeuSer---G 346
Db 438 CTGGTCACACCTGGCTCAAGGG-----CCTCCTCTCTGGCCTTAC 397
QY 346 lylLeuAlaTrpAlaValAlaMetIleAsnArgGlnGlnIleGlyGly-----ProA 363
Db 396 TGTCGCCCTGGAAGCTCCTC-----CAGTTTGGGGGCTGGGGAGTCCCCCT 352
QY 363 rgsrTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnProAlaCysP 383
Db 351 GTGGTCCCCCCTAGCAAGCAGTGTGAGGAGGAGGCCACTGGCTGTGTGCTGTGTGTG 292
QY 383 heileThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThrSer 401
Db 291 TGCTGACCCAGAGGACACACACTGCGCCAGGCGCTGGGT-----TGGAGCTCG 245
```

```
RESULT 14
US-10-750-185-55673
; Sequence 55673, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KEER, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 55673
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Bovine 19856880894405
US-10-750-185-55673

Alignment Scores:
Pred. No.: 16.2 Length: 1212
Score: 85.00 Matches: 41
Percent Similarity: 30.57% Conservative: 18
Best Local Similarity: 21.24% Mismatches: 62
Query Match: 3.67% Indels: 72
DB: Gaps: 8

US-10-602-219-12 (1-421) x US-10-750-185-55673 (1-1212)
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db 16 GCCTGTCTAGTGGCCCCCAGGAGGTGGGTGGCAGACTCC----- 54
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAla-----AspIle 232
Db 55 -----AGAAACACTGCATCTGGGGGCTCTGTGCTCCGCCCTCTCCTGTG 96
QY 233 AspAspSerTrpLysSerIleLysSerIleLeuAspTrpThrSerPheAsnGlnGluArg 252
Db 97 GACCAATGTGTGTTGATTGGGGTCTCTCAFTTCTGGCGCTCGTTCTTTGTATG--- 153
QY 253 IleValAspValAlaGlyProGlyGlyTrpAsn-----AspProAspMetLeuVal 269
Db 154 -----TGGNACAAGGGGGTTGACCTTGATTGCAAGTA 186
QY 270 Ile----- 270
Db 187 TTAGAACCCTCAGTTCGTTTAGTCTTTCTGGAGGAACCTCAAGAGAGGGCTTGTGTACA 246
QY 271 -----GlyAsnPheGlyLeuSerTrpAsnGlnValThrGlnMetAlaLeuTrp 287
Db 247 GCAAGCCTGGAGATGACAGGGCCCGAGCTGG-----GTCGTG 282
QY 288 AlalleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSerProGlnAla 307
Db 283 TCTCAGGGTGTCTGCTCTGCTCGGCTCCCTGGGTGGAGTTTGTGTTTCCGGCAGCC 342
QY 308 LysAlaLeuLeuGlnAspLys-----AspValIleAlalleAsn 320
Db 343 TCGCCCCGCTCAGGGCAGCAGGAGGGCCCTGGAGCTCTGTGATCTGTGTTTACCAGCT 402
QY 321 GlnAspProLeuGlyGlnGlyTrpGlnLeuArgGlnGlyAspAsn---PheGluVal 339
Db 403 GACAGCCCCCAAGAAAGGGGCCACCTCTCTCCATGGTTCTAGCCCTGGGAGCCCT 462
QY 340 TrpGluArgProLeuSerGlyLeuAlaTrpAlaValAla 352
```


Qy 173 -----TyrCysAspSerLeuGluAsnLeuAlaAspGlyTyr 184
Db 703 GTTTCAAAGGCTGGGTTTTCGCTCATATTCAACTCACACCCAAAGAGCGACCGTAC 762
Qy 185 LysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyrSerCysGluTrp 204
Db 763 AAGATGTG-----CTGTGGTCTATGATGGTGAATAT 795
Qy 205 ProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArgGlnTyrCysAsn 224
Db 796 GGGTTC-----TTTTTGGCCCAACGAGTATGGGTGATCAAGTCTGTGAC 843
Qy 225 HisTrpArgAsnPheAla 230
Db 844 GAGTTCCCTGGTTCTCT 861

Search completed: December 26, 2005, 16:10:01
Job time : 810 secs

Db 463 CCCTCTCTCCTGCTCCAGCTCTGTGTTGGGACCTGGCT 501
RESULT 15
US-11-073-626-4
; Sequence 4, Application US/11073626
; Publication No. US20050244926A1
; GENERAL INFORMATION:
; APPLICANT: KUROSAWA, KEIKO
; APPLICANT: HIROKAWA, KOZO
; APPLICANT: KAJIYAMA, NAOKI
; TITLE OF INVENTION: Novel fructoysl peptide oxidase
; FILE REFERENCE: 227590US0
; CURRENT APPLICATION NUMBER: US/11/073,626
; CURRENT FILING DATE: 2005-03-08
; PRIOR APPLICATION NUMBER: US/10/232,655
; PRIOR FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: JP 2001-266665
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: JP 2001-378151
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: JP 2002-228727
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 1314
; TYPE: DNA
; ORGANISM: Eupenicillium terrenum
US-11-073-626-4

Alignment Scores:
Pred. No.: 18.2 Length: 1314
Score: 85.00 Matches: 46
Percent Similarity: 33.19% Conservative: 29
Best Local Similarity: 20.35% Mismatches: 71
Query Match: 3.67% Indels: 80
DB: 9 Gaps: 12

US-10-602-219-12 (1-421) x US-11-073-626-4 (1-1314)

Qy 32 LeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrpLeuHisTrpGluArgPheMet 51
Db 343 CTCGATCGGGCATTTGGGCTGGAGAAGACGAACGTTTGGCTG----- 384
Qy 52 CysAsnLeuAspCysGlnGluProAspSerCysIleSerGluLysLeuPheMetGlu 71
Db 385 -----GAATCTGAAGATGAGATCCTCGCAAGAGCCCGAATTTACG 426
Qy 72 MetAlaGluLeuMetValSerGluGlyTrpLysAspAlaGlyTyrGluTyrLeuCysIle 91
Db 427 CGTGAACAAGTC-----AAGGGTGGAAAGGC-----TTATTTTGCACT 465
Qy 92 AspAspCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGln 111
Db 466 GATGGAGGCTGGCTTGTCTGCAGCAAG----- 492
Qy 112 ArgPheProHisGlyIleArgGlnLeuAlaAsnTyrValHisSerLysGlyLeuLysLeu 131
Db 493 -----GCTATCAATCGCATCGGAATTTTCTCCAGGACAAAGGTGTCAAGTTT 540
Qy 132 Gly-----IleTyrAlaAspValGlyAsnLysThr 141
Db 541 GCGTTTGGAGGTGCTGGACATTTTCAGCAACCTCTGTTGCGCGCTGATGGA---AAACT 597
Qy 142 CysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAlaAsp 161
Db 598 TGCATCGGACTTGAACTACACACGA-----ACCAAGTACTTTGCTGAC 642
Qy 162 -----TrpGlyValAspLeuLeuLysPhe---AspGlyCys 172
Db 643 AAGGTTGTCTTGGCTGGTGGTGGGATCCACCTTGGTGGATCTAGAGATCAGTGT 702

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 26, 2005, 12:40:36 ; Search time 890 Seconds
(without alignments)
3911.695 Million cell updates/sec

Title: US-10-602-219-12

Perfect score: 2314

Sequence: 1 MQLRNPGLHLCALALRFLA.....RLRSHINPTGTVLLQLENTM 421

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp

-Q=/cgn2_1/USPTO spool_p/US10602219/runat_23122005_151144_17253/app_query.fasta_1.583

-DB=Published Applications NA Main -QFMT=fastap -SUFFIX=p2n.rnpbm

-MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1

-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct

-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext

-HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000

-USER=US10602219 @CEN 1 1 1549 @runat_23122005_151144_17253 -NCPU=6 -ICPU=3

-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA Main:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*

2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*

3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*

4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*

5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*

6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*

7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*

8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*

9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*

10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	1266	3	US-09-993-059-11 Sequence 11, Appl
2	2314	100.0	1266	5	US-10-103-327-11 Sequence 11, Appl
3	2314	100.0	1266	7	US-10-602-219-11 Sequence 11, Appl
4	2314	100.0	1266	7	US-10-602-220-11 Sequence 11, Appl
5	2314	100.0	1266	8	US-10-851-388-11 Sequence 11, Appl
6	2314	100.0	1266	9	US-10-984-389-11 Sequence 11, Appl
7	2314	100.0	1278	3	US-09-993-059-7 Sequence 7, Appl

8	2314	100.0	1278	5	US-10-103-327-7	Sequence 7, Appl
9	2314	100.0	1278	7	US-10-602-219-7	Sequence 7, Appl
10	2314	100.0	1278	7	US-10-602-220-7	Sequence 7, Appl
11	2314	100.0	1278	8	US-10-851-388-7	Sequence 7, Appl
12	2314	100.0	1278	9	US-10-984-389-7	Sequence 7, Appl
13	2314	100.0	1284	3	US-09-993-059-13	Sequence 13, Appl
14	2314	100.0	1284	5	US-10-103-327-13	Sequence 13, Appl
15	2314	100.0	1284	7	US-10-602-219-13	Sequence 13, Appl
16	2314	100.0	1284	7	US-10-602-220-13	Sequence 13, Appl
17	2314	100.0	1284	8	US-10-851-388-13	Sequence 13, Appl
18	2314	100.0	1284	9	US-10-984-389-13	Sequence 13, Appl
19	2314	100.0	1290	3	US-09-993-059-3	Sequence 3, Appl
20	2314	100.0	1290	5	US-10-103-327-3	Sequence 3, Appl
21	2314	100.0	1290	7	US-10-602-219-3	Sequence 3, Appl
22	2314	100.0	1290	7	US-10-602-220-3	Sequence 3, Appl
23	2314	100.0	1290	7	US-10-411-037-67	Sequence 67, Appl
24	2314	100.0	1290	7	US-10-411-026-67	Sequence 67, Appl
25	2314	100.0	1290	7	US-10-410-962-67	Sequence 67, Appl
26	2314	100.0	1290	7	US-10-411-049-67	Sequence 67, Appl
27	2314	100.0	1290	7	US-10-410-930-67	Sequence 67, Appl
28	2314	100.0	1290	7	US-10-410-997-67	Sequence 67, Appl
29	2314	100.0	1290	7	US-10-411-012-67	Sequence 67, Appl
30	2314	100.0	1290	7	US-10-410-913-67	Sequence 67, Appl
31	2314	100.0	1290	8	US-10-851-388-3	Sequence 3, Appl
32	2314	100.0	1290	8	US-10-410-980-67	Sequence 67, Appl
33	2314	100.0	1290	9	US-10-410-897-67	Sequence 67, Appl
34	2314	100.0	1290	9	US-10-984-389-3	Sequence 3, Appl
35	2314	100.0	1296	3	US-09-993-059-9	Sequence 9, Appl
36	2314	100.0	1296	5	US-10-103-327-9	Sequence 9, Appl
37	2314	100.0	1296	7	US-10-602-219-9	Sequence 9, Appl
38	2314	100.0	1296	7	US-10-602-220-9	Sequence 9, Appl
39	2314	100.0	1296	8	US-10-851-388-9	Sequence 9, Appl
40	2314	100.0	1296	9	US-10-984-389-9	Sequence 9, Appl
41	2314	100.0	1306	9	US-10-480-790-5	Sequence 5, Appl
42	2314	100.0	1308	3	US-09-993-059-5	Sequence 5, Appl
43	2314	100.0	1308	5	US-10-103-327-5	Sequence 5, Appl
44	2314	100.0	1308	7	US-10-602-219-5	Sequence 5, Appl
45	2314	100.0	1308	7	US-10-602-220-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-11
; Sequence 11, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-09-993-059-11

Alignment Scores:	7,21e-278	Length:	1266
Pred. No.:	2314.00	Matches:	421
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	100.00%	Gaps:	0
DB:	3		


```
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAAAGGAGAGACTTCAGGAGACCTCAGCGCTTCCTCATGGGATTTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerIleGlyLeuGlyLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTACAGCAAGAGAGCTGAAGCTAGGGATTATGTCAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCTGGGAGTTTGGATACAGACATGATGATGCCAGACCTTGGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGlnAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGAAAAATTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTTCCTGGCCCTGAAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCGTGTAGTGGCCTCTTTATATGTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCATCACTCTGGCGAAATTTTCTGACATTGATGATTCCTCGAAAGATATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGAGCTGACATCTTTTAAACAGGAGAGAAATTTGTATGTTGCTGGACACGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GTTGGAAATGACCCAGATGATGTAGTGTGGCAACTTTGGCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTAAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGAGCTAAATGGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTACAGAGGGAGACAACCTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAAACCCGAGGAGATTGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTGCTTCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTCGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAAGCTAGGGTTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTGGCTTCAGCTAGAAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
```

RESULT 3

```
US-10-602-219-11
; Sequence 11, Application US/10602219
; Publication No. US20040016021A1
```

GENERAL INFORMATION:

```
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-11
```

Alignment Scores:

Pred. No.:	7,21e-278	Length:	1266
Score:	2314.00	Matches:	421
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-10-602-219-12 (1-421) x US-10-602-219-11 (1-1266)

Qy	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCGCTTCCTGGCC	60
Qy	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
Db	61	CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT	120
Qy	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
Db	121	ACCATGGGCTGGCTGACCTGGGAGCGCTTCAATGCAACCTTGATCTGCAGGAAGGCCA	180
Qy	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
Db	181	GATTCTGCATCAGTGAGAACTCTTCATGCAGATGGCAGAGCTCATGCTCTCAGAAGGC	240
Qy	81	TrpLysAspAlaGlyTyrGluTyrIleuCysIleAspAspCysTrpMetAlaProGlnArg	100
Db	241	TGGAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTGGATTGGCTCCCAAGA	300
Qy	101	AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
Db	301	GATTTCAGAAAGGAGAGACTTCAGGAGACCTTCAGGCTTCCTCATGGGATTTCGCCAGCTA	360

```
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyLeuValAlaAspValGlyAsnLys 140
Db |||
Db 361 GCTAATTATGTTACACGCAAGAGCTGAAGCTAGGAGTTATGACAGATTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db |||
Db 421 ACCTGCGCAGGCTTCCCTGGAGTTTGGATACTACGACATTCATGCTGCCAGACCTTTGCT 480
QY 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLysLeuLysLeu 180
Db |||
Db 481 GACTGGGAGTAGATCTGCTAAATTTTGATGGTTGTTACTGTGACAGATTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db |||
Db 541 CGAGATGGTTTAAGCACATGCTCTTGGCCCTGATAGACTGGCAGAGCAATTTGTGTAC 600
QY 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluLeuArg 220
Db |||
Db 601 TCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTyrPheGlnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db |||
Db 661 CAGTACTGCAATCACCTGGCGAAATTTTGCTGCATTTGATGATTCCTGGGAAAGTATAAG 720
QY 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db |||
Db 721 AGTATCTTGATCTGGACATCTTTTACAGAGAGAGAAATTTGTGATGTTGCTGGACGAGG 780
QY 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db |||
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTyrPheAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db |||
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db |||
Db 901 CGACACATCAGCCCTCAGCCCAAGCTCTCTTCAGGATPAGGACGTAAATGCCATCAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db |||
Db 961 CAGGACCCCTTGGCAAGCAAGGTTACCAGCTTAGACAGGAGACAACCTTTCAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTyrPheAlaValAlaMetIleAsnArgGlnIleGly 360
Db |||
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATATATAAACCGCAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db |||
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCCCTGTATCTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db |||
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db |||
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTCAGCTAGAAAACACA 1260
QY 421 Met 421
Db |||
Db 1261 ATG 1263
```

RESULT 4

```
US-10-602-220-11
; Sequence 11, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
```

```
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-220-11
```

Alignment Scores:

Pred. No.:	7,21e-278	Length:	1266
Score:	2314.00	Matches:	421
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-10-602-219-12 (1-421) x US-10-602-220-11 (1-1266)

QY	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCGCTTCCTGGCC	60
QY	21	LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
Db	61	CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGGATTGGCAAGGACGCT	120
QY	41	ThrMetGlyTyrLeuHisTyrProGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
Db	121	ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTCAACCTTGACTGCCAGAGAGCCA	180
QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
Db	181	GATTCCTGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGC	240
QY	81	TyrLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrMetAlaProGlnArg	100
Db	241	TGGAAGGATGCAGGTTATGAGTACCTTCGCAATGATGACTGTGTGATGGCTCCCCAAGA	300
QY	101	AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
Db	301	GATTCAGAAGGCAGACTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGTA	360
QY	121	AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys	140
Db	361	GCTAATTATGTTACACGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAAATAAA	420

```
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTCGCAGGCTCCCTCGGGAGTTTGGATACACGATGATGCGCCAGACCTTTCCT 480
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTTGTACTGTGACAGATTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTGCGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTAGTGGCCCTCTTTATATGCGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCATCATCTGGCGAAATTTGCTGCATTTGATGATCTCTGGAAAGTATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAACACAGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGAAATGACCCAGATGATGTAGTGTATGCTGCTGCTCTTTTATTCATGTCATAAGACCTC 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCATAAGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGAGTAAATGGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACACCTCTCTCAGCTTAGCTGAGCTGTAGCTATGATAAACCAGGAGAGATGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGACGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGGAAGTAGGTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuAsnThr 420
Db 1201 TCAGGTAAAGAAAGTCATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAAACACA 1260
QY 421 Met 421
Db 1261 ATG 1263
```

RESULT 5

```
US-10-851-388-11
; Sequence 11, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Tomo H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; CURRENT FILING DATE: 2004-05-21
```

```
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-851-388-11
Alignment Scores:
Pred. No.: 7,21e-278 Length: 1266
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0
```

US-10-602-219-12 (1-421) x US-10-851-388-11 (1-1266)

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTTCCTTCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCATCTGGACATGATGATTTGGCAAGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCGAGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGATCAGTGAGAGCTCTTCATGAGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGATGAGAGTTATGATGATCTCTGCATTTGATGATCTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAGGACAGCTTCAGGAGAGCCCTCAGGCGCTTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATTATGTTACAGCAAGAGCTGAAGTAGGGATTTATGACAGATGTTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGGAGGCTTCCCTGGGAGTTTGGATACACGACATTTGATGCGCCAGACCTTTCCT 480
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTTGTACTGTGACAGATTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTGCGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTAGTGGCCCTCTTTATATGCGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCATCATCTGGCGAAATTTGCTGCATTTGATGATCTCTGGAAAGTATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
```

Db 721 AGTATCTTGGACTGGACATCTTTTAAACCAAGGAGAGAAATGTTGATGTTGCTGGACACGGG 780
Qy 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGATGACCAAGATATGATGATGGCACTTTGGCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTTATTATTCATGTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCGCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAsnAspPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGACCAAGCTTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAACCGGACGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 CGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGlnTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCCTCTGTGAAGAGAACTAGGGTTCATGAAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAAAGTCACATNAATCCACAGGCACCTGTTTTCCTCAGCTAGAAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263

RESULT 6

US-10-984-389-11
; Sequence 11, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURBEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; PRIOR FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-984-389-11

Alignment Scores:

Pred. No.:	7,21e-278	Length:	1266
Score:	2314.00	Matches:	421
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-602-219-12 (1-421) x US-10-984-389-11 (1-1266)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACACTGGGCTGGCGCTTGGCTTCGCTTCTCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluLeuPro 60
Db 121 ACCATGGGCTCGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGGCAGGAAGGCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGGAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAAGGCAGACTTCAGGCGACCTCAGCGCTTTCCTCATGGGATTGGCCAGTA 360
Qy 121 AlaSerTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTTACAGCAAGAGACTGAAGTAGGATTTATGCAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACAGACATTCATGCTCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATCGTTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTATAGCACATGTCCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCCTCTTATATATGGGCCCTTTCAAAAGCCCAATTATACAGAAATCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCCTGGCGAAATTTTGTCTGACATTTGATGATTCCTGGAAAGTATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACCAAGAGAGAAATGTTGATGTTGCTGGACACGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTATTGGCAACTTTGGCACTTCAGCTCAGGAATCAGCA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTTATTCATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCGCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnAspPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGCTACCGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATATAACCGGACGAGATTGGT 1080

QY 361 GlyProArgSerThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTATACCATCGAGTCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheThrGluTrpThr 400
 Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGAGAGCTAGGTTCTATGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuAsnThr 420
 Db 1201 TCAAGGTTAAGAAGTACATATAATCCACAGGCACTGTTTGGCTTCAGCTAGAAAAACACA 1260
 QY 421 Met 421
 Db 1261 ATG 1263

RESULT 7

US-09-993-059-7
 ; Sequence 7, Application US/09993059
 ; Publication No. US20020088024A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GARGER, Stephen A.
 ; APPLICANT: TURPEN, Thomas H.
 ; APPLICANT: KUMAGAI, Monto H.
 ; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
 ; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
 ; FILE REFERENCE: 008010087CFUS06
 ; CURRENT APPLICATION NUMBER: US/09/993,059
 ; CURRENT FILING DATE: 2001-11-13
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 1278
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1275)
 US-09-993-059-7

Alignment Scores:
 Pred. No.: 7,31e-278 Length: 1278
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0

US-10-602-219-12 (1-421) x US-09-993-059-7 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGGCTGGGCTTCCTCTCTGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlnLeuAlaArgThrPro 40
 Db 61 CTCGTTCTCTGGACATCCCTGGGCTAGAGCACTGGCAATGGATGGCAAGGAGCGCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGGCTGGCTGCACTGGGAGGCGCTTCATGTGCAACTTGGACTGCCAGGAAGGCCA 180
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCTGTCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAGGATGCAAGTTATGATGACCTCTGATGATGATGATGATGATGATGATGATGATGAT 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTTCAGAGGCGAGACTTCAGGCGAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

QY 121 AlaIleThrValHisSerLysGlyLeuLysLeuGlyIleThrAlaAspValGlyAsnLys 140
 Db 361 GCTAATATGTTTACAGCAAGAGCTGAAGCTAGGATTTATGCGAGATGTTTGGAAATAAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTTGGATCTACGACATTTGATGCCAGACCTTTGCT 480
 QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGATCTGTAAATTTGATGGTTTACTGTGACAGTTTGGAAATTTG 540
 QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
 Db 541 GCAGATGGTTATAAGCACATGCTTGGCCCTGTAATAGGACTGGCAGAAGCATTTGTGTAC 600
 QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
 Db 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
 QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTGCATCACTGGCGAAATTTGCTGACATTTGATGATTTCTGGAAAGTATAAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACCCAGGAGAAATTTGATGATTTGCTGGACCGAG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 781 GGTGGAAATGACCCAGATATGTTAGTGTGGCAACTTTGGCCCTCAGCTGGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACTCAGATGGGCTCTGGGCTATCATGCTGCTCTCTTTTATTTATGCTTAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGGCAAGAGGTTACAGCTTTAGACAGGAGAGCAACTTTGAAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCCGCGAGGAGATTGGT 1080
 QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCGAGTCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheThrGluTrpThr 400
 Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGAGAGCTAGGTTCTATGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 Db 1201 TCAAGGTTAAGAAGTACATATAATCCACAGGCACTGTTTGGCTTCAGCTAGAAAAACACA 1260
 QY 421 Met 421
 Db 1261 ATG 1263

RESULT 8

US-10-103-327-7
 ; Sequence 7, Application US/10103327
 ; Publication No. US20030106095A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GARGER, Stephen A.
 ; APPLICANT: TURPEN, Thomas H.
 ; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-10-103-327-7

Alignment Scores:
Pred. No.: 7,31e-278 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-602-219-12 (1-421) x US-10-103-327-7 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCTCGCTTCTGGGCC 60
QY 21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGACACTGGCAATGGATTGGCAAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCATCTGGGAGCGCTTCATGTGCACCTTGACTGCCAGGAAGGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
QY 81 TrpLysAspAlaGlyTyrGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAGGATGCAAGGTATAGTACCTCTGCATTTGATGACTGTGGTGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGACTTCAGGAGACCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGAGACTGAAGTAGTGGGATTTATGAGATGTTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGTGATATAGTATGATGATGATGATGATGATGATG 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAAATTGATGTTGATGTTGATGTTGATGTTGATGTTGATG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCTGTGAGTGGCCTCTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240

RESULT 9

US-10-602-219-7
; Sequence 7, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31

; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-7

Alignment Scores:
Pred. No.: 7,31e-278 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00 Conservative: 0
Best Local Similarity: 100.00 Mismatches: 0
Query Match: 100.00 Indels: 0
DB: 7 Gaps: 0

US-10-602-219-12 (1-421) x US-10-602-219-7 (1-1278)

QY 1 MetGlnLeuArgAspProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGGCTTGGCTTGGCTTCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCTGGGACATCTCCGGCTAGAGCCTGGACATGATGGATGGCAAGAGGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGGCTGGCAGGAGCCA 180
QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTGAAGACTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGAAGAGATGCAAGTTATGAGTACCTCTGCATGTATGATGCTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheHisGlyIleArgGlnLeu 120
DB 301 GATTCAGAAGGCAGACTTCAGGACAGCCCTCAGCGCTTCTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 361 GCTAATATGTTTCACAGCAAGAGACTGAAGCTAGGAGTTATGCAAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATGATGATGATGATGATGATGATGATGATGAT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTGAATCTGCTAAATTTGATGTTGATGTTGATGTTGATGTTGATGTTGATG 540
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
DB 541 CGAGATGGTTATAAGCACATGCTCTGGCCCTGATAGGACTGGCAGAGCATTTGTTGATC 600
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
DB 601 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTTTCAAAAGCCCAATATATACAGAAATCCGA 660
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCATACCTGCGCAATTTTGTGATGATGATGATGATGATGATGATGATGATGAT 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTATGTTGCTGGACCAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
DB 781 GGTGGATGACCCAGATATGTTAGTATTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACTCAGATGGGCCCTCTGGGCTATCATGCTCTCTTTTATTCTATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACATCATGCCCTCAAGCCCAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGAGCAACTTTGGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCCGAGGAGATTGGT 1080
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GCACCTCTCTTTATACCATGCAGTTGCTTCCCTGGTAAAGGAGTGGCCTGTATCTCT 1140
QY 381 AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
DB 1141 GCCTGCTTCATCACACAGCTCTCTCTCTGTAAGAGGAGTAGGCTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1201 TCAAGTTTAAAGAGTCACATAATCCACAGGCACTGTTTGTCTTCAGTAGAAAAACA 1260
QY 421 Met 421
DB 1261 ATG 1263

RESULT 10
US-10-602-220-7
; Sequence 7, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

```
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-220-7

Alignment Scores:
Pred. No.: 7, 31e-278 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-219-12 (1-421) x US-10-602-220-7 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTTGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGGTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGAGCGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCAATGTGCAACCTTTGACTGCCAGGAGAGCCA 180

QY 61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAGGTATCAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTACAGCAAGAGACTGAAGCTAGGAGATTATGCAGATGTTGGAATAAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGAGTTTGGATACACGATGATGATGCCAGACCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTACACATGTTGGAAAAATTG 540

QY 181 AlaAspGlyTyrIleHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTTGGCCCTGAATAGGACTGGCAGAACATTTGTATC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGCCCAATATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCTGACATTTGATGATTCCTGGAAAAAGTAAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValalaglyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGATGTTCTGACAGGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTTCTGGCC 60
```

```
Db 781 GGTGGAAATGACCCAGATATGTTAGTTAGTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
QY ValThrGlnMetAlaLeuTrpAlaIleMetAlaIleProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTGGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGGGTACAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCCGCGAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGAAAGCTAGGCTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTTGCTTCAGTAGAAAAACACA 1260
QY 421 Met 421
Db 1261 ATG 1263

RESULT 11
US-10-851-388-7
; Sequence 7, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUNAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; PRIOR FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-10-851-388-7

Alignment Scores:
Pred. No.: 7, 31e-278 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

US-10-602-219-12 (1-421) x US-10-851-388-7 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTTCTGGCC 60
```

```
QY 21 LeuValSerTrpAspProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGATGGCAAGGAGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCATCGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAGCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGAGATGCAGGTATGATGATCTGCTGATGATGATGATGATGATGATGATGATGAT 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAAGGAGAGCTTCAGGAGAGCTTCAGGAGAGCTTCAGGAGAGCTTCAGGAGAG 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGAGACTGAGCTAGGATTTATGAGATGTTGGAAATATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTCGCGAGGCTCCCTGGGAGTTTGGATACACTAGCATTGATGATGATGATGATGATGAT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLysLeuAsnLeu 180
Db 481 CACTCGGAGTAGATCTGCTAAATTTTGTATGTTTGTATGTTTGTATGTTTGTATGTTTGT 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 CGAGATGGTTATAGACACATGCTTGGCCCTGAAATAGGACTGGCAGAGCAATGTTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGCCCTTTATATGTTGTCCTTTTCAAGGCCCAATATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTCGGCGAAATTTTGTGCACTTGTATGATTCCTCGAAAGATATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTATGGCACTTTGGCTCAGCTGGAAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGAGCGTAATTCGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCCAGCAGGATACAGCTTAGACAGGAGAGCAACTTTGAAGTGTTG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluLeGly 360
Db 1021 GAACGACCTCTCTCAGCTTAGCTGGGCTGATGATGATGATGATGATGATGATGATGATGAT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGCCCTGTATCTCT 1140
```

```
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTTGAAAAAGGAGCTTAGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGTTTAAAGTAGTCATATAATCCCAAGGAGCTGTTTGTCTCAGTAGAANAACACA 1260
QY 421 Met 421
Db 1261 ATG 1263

RESULT 12
US-10-984-389-7
; Sequence 7, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Tomohiko H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; CURRENT FILING DATE: 2004-11-08
; PRIOR FILING DATE: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-10-984-389-7

Alignment Scores:
Pred. No.: 7,31e-278 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-602-219-12 (1-421) x US-10-984-389-7 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTGCTTCTGCGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGATGGCAAGGAGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCATCGGAGGCTTCATGTCACCTTGACTGCCAGGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAGCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGAGATGAGGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAAGGAGAGCTTCAGGAGAGCTTCAGGAGAGCTTCAGGAGAGCTTCAGGAGAG 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
```

```
Db 361 GCTAATATGTTACAGCAAGAGCTAGAGCTAGGATTATGCGAGATGTTGCAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTCCCTGGGAGTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
Qy 161 AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGAATCTGTAATAATTTGATGTTGTTACTGTACAGATTGGAAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 CGAGATGGTTATAAGCACATGCTCTTGGCCCTGATAGGACTGGCAGAACATTTGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCANATCCTGGCGAATTTTGTGCAATGATGATTCCTGGAAAAGTATAAG 720
Qy 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAACACAGGAGAGAAATTTGTTGTTGCTGCACAGGG 780
Qy 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTCGAATTTGGCCCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCCTCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGCAAGCAAGGTTACCACTTACAGCTTAGACAGGGAGACAACCTTTGAAGTGTG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTACCTGGCTGAGCTATGATATAACCGGACGAGAGATTGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGTCTCATCACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeuGluAsnThr 420
Db 1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
```

RESULT 13

```
US-09-993-059-13
; Sequence 13, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
```

```
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ... (1284)
US-09-993-059-13

Alignment Scores: 7.36e-278 Length: 1284
Pred. No.: 2314.00 Matches: 421
Score: 2314.00 Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00% Gaps: 0
DB: 3
```

US-10-602-219-12 (1-421) x US-09-993-059-13 (1-1284)

```
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGGCGCTTGGCTTCGCTTCCTGGCC 60
Qy 21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGluGluPro 60
Db 121 ACCATGGGCTGGCTGCATCTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGAAAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrMetAlaProGlnArg 100
Db 241 TGGAAAGGATGAGGTTATGAGTACCTCTGCATTCATGACTGTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAGGAGAGATTCAGGACCTTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTACAGCAAGGACTGAAGCTAGGATTTATGCAAGATGTTGGAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTCCCTGGGAGTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
Qy 161 AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGTTGTTTGTGACAGTTTGGAAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGATAGGACTGGCAGAACATTTGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCANATCCTGGCGAATTTTGTGCAATGATGATTCCTGGAAAAGTATAAG 720
Qy 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
```

```
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
Qy 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTATGGCACTTTGGCTCAGCTGGNAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaIleProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaIleValLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTGGGCAAGCAGGATCCAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGCAGGATGGT 1080
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGACTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGAGAGTGGGTTCTATGAAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTTAAGAGTACACATAAATCCACAGGCACTGTTTCTTTCAGCTAGAAAAACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
```

RESULT 14

```
US-10-103-327-13
; Sequence 13, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KOMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-10-103-327-13
```

Alignment Scores:

```
Pred. No.: 7,36e-278 Length: 1284
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
```

```
US-10-602-219-12 (1-421) x US-10-103-327-13 (1-1284)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGCCTTCGCTTCCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGCTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGAGGCTTCATGTGCAACCTTGACTGTCGAGGAAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGTCATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGGAGGAGACTTCAGGAGACCTCAGGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaSerTrpValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGGAATTTATGCAGATGTTGGAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATCTACGACATTTGATGCCAGACCTTTGT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTGTGTTACTGTGACAGCTTTGGAAAATTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTATAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrTrpGluIleArg 220
Db 601 TCCTGTAGTGGCCCTCTTTATATGTGGCCCTTTTCAAAAGCCCAATATATACAGAAAT 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATAAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaIleProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTGGGCAAGCAGGATCCAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGCAGGATGGT 1080
```


Qy	361	GlyProArgSer	TyrThr	LeuAlaVal	AlaSer	LeuGly	LeuGly	ValAla	CysAsn	Pro	380	
Db	1081	GGACCTCGCT	TATAC	GCATCG	AGTTGC	TTCCTGGG	TAAAG	GAGTGG	CTGTAAT	TCCT	1140	
Qy	381	AlaCysPhe	LeuThr	GlnLeu	ProVal	LeuArg	GlyLeu	GlyPhe	TyrGlu	TyrThr	400	
Db	1141	GCCTGCTT	TCATCAC	ACAGCT	CTCCCTCGT	GAAAG	GAAGCTAG	GGTTC	TATGAT	TGGACT	1200	
Qy	401	SerArgLeu	ArgSer	HisLeu	AsnPro	ThrGly	ThrVal	LeuLeu	GlnLeu	GluAsn	Thr	420
Db	1201	TCAGGTTA	GAGTGC	ATAAAT	CCACAGG	CACTGTTT	TGCTT	CAGCTAG	GAACACA		1260	
Qy	421	Met	421									
Db	1261	ATG	1263									

RESULT 15

```

US-10-602-219-13
; Sequence 13, Application US/10602219
; Publication NO. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: LSHC-0087-CPD09B

```

Alignment Scores:		
Pred. No.:	7,366-278	1284
Score:	2314.00	Matches: 421
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	100.00%	Indels: 0
DB:	7	Gaps: 0

US-10-602-219-12 (1-421) x US-10-602-219-13 (1-1284)

[illegible]

QY	381	AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr	400
Db	1141	GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAGCTAGGGTTCTATGAATGACT	1200
QY	401	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr	420
Db	1201	TCAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAACACA	1260
QY	421	Met	421
Db	1261	ATG	1263

Search completed: December 26, 2005, 15:56:37
 Job time : 908 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 26, 2005, 10:34:12 ; Search time 4085 Seconds
(without alignments)
4821.876 Million cell updates/sec

Title: US-10-602-219-12
Perfect score: 2314
Sequence: 1 MOLRNPEHLGALALRFLA.....RLRSHINPTGVLLQLENTM 421

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10602219/runat_23122005_151142_17149/app_query.fasta_1.583
-DB=EST -QPM=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602219 @CEN 1 1 8010 @runat_23122005_151142_17149 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT_DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	1277	4	CR605654 full-leng
2	2309	99.8	1266	4	CR607242 full-leng
3	2289	98.9	1253	4	CR617861 full-leng
4	2104	90.9	1290	10	AY408540 Homo sapi
5	2055	88.8	1290	10	AY408541 Pan trogl
6	1811	78.3	1376	4	AK054547 Mus muscu
7	1811	78.3	2962	4	AK040081 Mus muscu

8	1777	76.8	1086	1	AL554978	AL554978
9	1699	73.4	1014	5	BX354096	BX354096
10	1691.5	73.1	999	1	AL552630	AL552630
11	1667.5	72.1	1067	3	BM564282	BM564282
12	1659.5	71.7	1005	3	BQ062192	AGENCOURT
13	1649	71.3	922	7	CO645623	ILLUMIGEN
14	1648.5	71.2	1026	1	AL577581	AL577581
15	1632	70.5	1296	10	AY408542	Mus muscu
16	1627.5	70.3	1133	7	CO645672	ILLUMIGEN
17	1589	68.7	920	7	CO645464	ILLUMIGEN
18	1578	68.2	923	7	CT005156	CT005156
19	1574	68.0	984	7	CO646251	ILLUMIGEN
20	1556.5	67.3	976	5	BQ956043	AGENCOURT
21	1527	66.0	898	5	BUI91867	AGENCOURT
22	1519	65.6	922	5	BUI540848	AGENCOURT
23	1515.5	65.5	937	5	BQ934640	AGENCOURT
24	1492.5	64.5	852	6	CA454083	AGENCOURT
25	1465	63.3	888	5	BUI54569	AGENCOURT
26	1458	63.0	958	6	CA487415	AGENCOURT
27	1452	62.7	871	6	CA454143	AGENCOURT
28	1443	62.4	801	5	BUI596617	AGENCOURT
29	1429	61.8	1025	3	BM450649	AGENCOURT
30	1425.5	61.6	849	6	CA487531	AGENCOURT
31	1422.5	61.5	879	2	BG824387	AGENCOURT
32	1412.5	61.0	927	2	BE622583	AGENCOURT
33	1411.5	61.0	1071	1	AL575861	AGENCOURT
34	1408.5	60.9	871	3	BQ225444	AGENCOURT
35	1406	60.8	908	2	BE379425	AGENCOURT
36	1404	60.7	796	8	DR156332	HESC2 64
37	1402.5	60.6	921	6	CA488935	AGENCOURT
38	1397	60.4	845	5	CR983967	CR983967
39	1384	59.8	813	5	BX344841	EX344841
40	1382	59.7	782	2	BI224248	BI224248
41	1371.5	59.3	1066	2	BG824323	AGENCOURT
42	1365	59.0	749	5	BX344842	EX344842
43	1360	58.8	908	6	CD385342	AGENCOURT
44	1326	57.3	886	3	BP435628	BP435628
45	1303	56.3	767	3	BI753664	BI753664

ALIGNMENTS

RESULT 1
CR605654
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK

CR605654 1277 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CS0DB008YB11 of Neuroblastoma Cot
10-normalized of Homo sapiens (human).
CR605654 GI:50486461
CR605654 HTC; CNSLT_CDNA.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1277)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1277)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequenage :
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
- Web: www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CR605654 1277 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CS0DB008YB11 of Neuroblastoma Cot
10-normalized of Homo sapiens (human).
CR605654 GI:50486461
CR605654 HTC; CNSLT_CDNA.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1277)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1277)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequenage :
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
- Web: www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES
Source

Location/Qualifiers
1. .1277
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D067YJ01"
/tissue_type="Neuroblastoma Cot 10-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:

Pred. No.: 6.49e-243 Length: 1277
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-602-219-12 (1-421) x CR605654 (1-1277)

QY 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 10 ATGAGCTGAGGAACCCAGAACTACATCTGGGCTGGGCTGGGCTTGGCTTCTCGGCC 69
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 70 CTCGTTCTTGGACATCTCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGGAGCGCT 129
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 130 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTTGACTGCCAGGAAGGCCA 189
QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 190 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 249
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 250 TGAAGGATGCAGTTATGAGTACCTCTGCATTTGATGATGCTGGTGGTCTCCCAAGA 309
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 310 GATTCAGAAGGCAGACTTCAGGCAGACCTCAGGCTTCTCCTCATGGGATTCGCCAGCTA 369
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 370 GCTAATTATGTTTACACAGCAAGGACTGAAAGCTAGGGAATTTATGCAGATGTTGGAATAA 429
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 430 ACTGCGCAGGCTTCTCTGGAGTTTGGATATCTAGCATTTGATGCCAGACCTTTGCT 489
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 490 GACTGGGGAGTAGATCTGCTAAATTTTGATGGTTTACTGTGACAGCTTTGGAAATTTG 549
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 550 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAATAGACTGGCAGAACATTTGTTAC 609
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 610 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 669
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 670 CAGTACTGCAATCACTGGCGAAATTTTGTGCATTTGATGATGATTCCTGGAAAGTATAAG 729
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValIaGlyProGly 260
DB 730 AGTATCTTGGACTGGACATCTTTTAAACCAAGGAGAAATTTGTGATGTTCTGGACAGGG 789
QY 261 GlyTrpAsnAppProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280

DB 790 GGTGGATGACCCAGATATGTTAGTGATTGGCACTTTGGCCCTCAGCTGGAATCAGCAA 849
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaIaProLeuPheMetSerAsnAspLeu 300
DB 850 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTTATTTCATGCTCTAATGACCTC 909
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 910 CGACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGAGCTAATTTGCCATCAAT 969
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 970 CAGGACCCCTTGGGCAAGAGGTACAGCTTAGACAGGAGACACTTTTGAAGTGTGG 1029
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1030 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGCAGAGATTGGT 1089
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1090 GGACCTCGCTCTTATACCATCGAGTTCCTCCCTGGGTAAGGAGTGGCTGTGTAATCT 1149
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1150 GCCTGCTTCATCACACAGCTCCTCCTCTGTAAGGAAGTAGGCTTCTATGAATGGACT 1209
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1210 TCAAGTTTAAAGTAGTCATAAATCCACAGGCACTGTTTGTCTTTCAGTAAAAATACA 1269
QY 421 Met 421
DB 1270 ATG 1272
RESULT 2
CR607242
LOCUS full-length cDNA clone CS0D1067YJ01 of Placenta Cot 25-normalized
DEFINITION of Homo sapiens (human).
ACCESSION CR607242
VERSION CR607242.1 GI:50488049
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1266)
AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1266)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1. .1266
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1067YJ01"
/tissue_type="Placenta Cot 25-normalized"

/plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:
 Pred. No.: 2,27e-242 Length: 1266
 Score: 2309.00 Matches: 420
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.78% Indels: 0
 DB: 4 Gaps: 0

US-10-602-219-12 (1-421) x CR607242 (1-1266)

QY 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 DB 7 ATGACGCTGAGGAACCCAGAACTACATCTGGGCTCGGCGCTTGGCTTCCTCGGCC 66

QY 21 LeuValSerTrpAspIleProGlyValaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 DB 67 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCAGCTGGCAATGGATTGGCAAGGACGCT 126

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 DB 127 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCA 186

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 DB 187 GATTCCTGCATCAGTGAGAAGCTTTCATGTGAGATGGCAGAGCTCATGGTCTCAGAAGGC 246

QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 DB 247 TGAAGGATGACAGGTATGAGTACCTCTGCAATTTGATGATGCTGGTGGCTCCCAAGA 306

QY 101 AspSerGluGlyValArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 DB 307 GATTCAGAGGAGAGACTTCAGCAGACCTCAGCGCTTCTCTCATGGGATTCGCAGGTA 366

QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
 DB 367 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGATTTATGCAAGATGTTGGAAATAA 426

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
 DB 427 ACCTCGCAGGCTCCCTCGGGAGTTTGGATACACTACGACATGTATGCCACAGCTTCT 486

QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 DB 487 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 546

QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
 DB 547 GCAGATGGTTATAGACACATGCTCTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTA 606

QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
 DB 607 TCTGTGAGTGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 666

QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 DB 667 CAGTACTGCAATCACTGGCGAAATTTGCTGACATTGATGATCTCTGAAAGATATAAG 726

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 727 AGTATCTGGACTGCACATCTTTTAACCCAGGAGAGAAATTTGATGTTGCTGGACAGGG 786

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
 DB 787 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 846

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 DB 847 GTAACCTCAGATGGCCCTCTGGGCTCATGCGCTGCTCTCTTTATTCTGCTTAATGACCTC 906

QY 301 ArgHisIleSerProGlnAlaLysAlaIleuLeuGlnAspLysAspValIleAlaIleAsn 320
 DB 907 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCATCAAT 966

QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnGluValTrp 340
 DB 967 CAGGACCCCTTGGCAAGCAGGTTACGAGTTAGACAGGAGAGCACTTTGAGTGTGG 1026

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 DB 1027 GAAGGACCTCTCTCAGGCTTAGCCTGGCTAGCTATGATAAAACCGCAGGAGATTGGT 1086

QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 DB 1087 GGACCTCGCTTATACATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAATCCT 1146

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
 DB 1147 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAGGAGAGTGGGTTCTATGATGGACT 1206

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 DB 1207 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAAAATACA 1266

RESULT 3

CR617861 1253 bp mRNA linear HTC 21-JUL-2004

LOCUS full-length cDNA clone CS0DK009YE12 of HeLa cells Cot 25-normalized

DEFINITION of Homo sapiens (human).

ACCESSION CR617861

VERSION CR617861.1 GI:50498668

KEYWORDS HTC; CNSUT_CDNA.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1253)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue
 REFERENCE 2 (bases 1 to 1253)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail: seque@genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES
 Location/Qualifiers
 1..1253
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DK009YE12"
 /tissue_type="HeLa cells Cot 25-normalized"
 /plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:
 Pred. No.: 3,53e-240 Length: 1253
 Score: 2289.00 Matches: 416
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 98.92% Indels: 0
 DB: 4 Gaps: 0

US-10-602-219-12 (1-421) x CR617861 (1-1253)

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 6 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTCGGCC 65
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 66 CTCGTTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGCGCT 125
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 126 ACCATGGCTGGCTGCAGCTGGAGCGCTTCATGTGCACCTTGACTGCCAGGAAGGCCA 185
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 186 GATTCTCGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 245
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 246 TGGAAGGATGCAGGTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCCAAGA 305
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 306 GATTTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGTCA 365
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 366 GCTAAATTATGTTACACAGCAAGGACTGAAGCTAGGAGATTATGCAGATGTTGGAATAAA 425
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 426 ACCTGGCAGGCTTCCTGGAGTTTGGTACTACGACATGTATGATGCCAGACCTTTGCT 485
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 486 GACTGGGGAGTAGACTCTGCTAAATTTTGATGGTTGTTACTGTACACATGTTGGAAAATTG 545
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 546 GCAGATGGTTATGAACACATGCTCTGGCCCTGGAATAGGACTGGCAGAACATTTGTATC 605
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 606 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCA 665
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 666 CAGTACTGCAATCACCTGGCGAAATTTTGCTGACATTTGATGATTCCTCGAAAAGTATAAG 725
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValalaglyProGly 260
Db 726 AGTATCTTGACATGGACATCTTTTAACCAAGGAGAGAAATTTGTATGTTGCTGGACAGGG 785
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 786 GGTGTGAATGACCCAGATATGTAGTATGGCAACTTTGGCTCAGCTGGATCAGCA 845
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
Db 846 GTAACCTCAGATGGCCCTCTGGGGTATCATGGCTGCTCCTTTATTCATGTCTAATGACCTC 905
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 906 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCCTCAAT 965
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 966 CAGGACCCCTTGGCAGCAGCAGGGTACAGCTTAGACAGGGAGACAACTTTGAGGTGG 1025
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1026 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTTATGATAAACCGGAGGAGATTGCT 1085
```

```
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1086 GGACCTCGCTCTTATACCATGCAGTTGCTTCCCTGGGTAAGGATGGCCCTGTAATCCT 1145
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1146 GCCTGCTTTCATCACACAGCTCTCCCTGTGAAAAGGAAGTAGGTTCTATGAATGGACT 1205
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGln 416
Db 1206 TCAAGTTAAGAAGTCACATAATCCACAGCACTGTTTGTCTTCAG 1253

RESULT 4
LOCUS AY408540 1290 bp DNA linear GSS 15-DEC-2003
DEFINITION Homo sapiens GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY408540
VERSION AY408540.1 GI:39764511
KEYWORDS GSS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1290)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Inferred nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
JOURNAL PUBMED 14671302
REFERENCE 2 (bases 1 to 1290)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
location/Qualifiers
1..1290
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1_>1290
/gene="GLA"
/locus_tag="HCM3258"

gene
Alignment Scores:
Pred. No.: 7,95e-220 Length: 1290
Score: 2104.00 Matches: 390
Percent Similarity: 92.64% Conservative: 0
Best Local Similarity: 92.64% Mismatches: 31
Query Match: 90.92% Indels: 0
DB: 10 Gaps: 0

US-10-602-219-12 (1-421) x AY408540 (1-1290)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTCGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGCGCT 120
```

```
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAGGTATTGAGTACCTCTGCATTGTGATGCTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGCAGCTTCAGGAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyValHisSerLysGlyLeuLysLeuGlyIleTyAlaAspValGlnAsnLys 140
Db 361 GCTAATTATGTTACACAGCAAGAGCTGAAGCTAGGAGTTATGCGATGTTGGAAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACTCGGCAGGCTCCCTGGAGTTTGGATACCTACGACATTGATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyCysAspSerLeuGlnAsnLeu 180
Db 481 GACTGGGAGTAGACTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
QY 181 AlaAspGlyTyLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTy 200
Db 541 CGAGATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600
QY 201 SerCysGluTrpProLeuTyMetTrpProPheGlnLysProAsnTyThrGluIleArg 220
Db 601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660
QY 221 GlnTyCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCAGCTGGCGAAATTTGTGTCATTTGATGATTCCTGGGAAAGTATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTCGACTGGGACATCTTTTAAACAGGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTAGTATGGCACTTTGGCTCAGCTGGGAATCAGCA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGAGCGTAATTCGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGTTACCGCTTAGCAGGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGCTTAGCTGGCTGGCTGTAGCTATGATATAACCGGAGGAGATTGCT 1080
QY 361 GlyProArgSerTyThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
QY 381 AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTyTrpGluTrpThr 400
Db 1141 GCTGTCTTATCACAGCTCCCTCCCTGTGAAAGGAAGCTAGGGTCTTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
```

```
Db 1201 TCAGAGTTAAGAGTACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAATACA 1260
QY 421 Met 421
Db 1261 ATG 1263

RESULT 5
LOCUS AY408541 1290 bp DNA linear GSS 15-DEC-2003
DEFINITION Pan troglodytes GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY408541
VERSION AY408541.1 GI:39764512
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source 1..1290
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1290
/gene="GLA"
/locus_tag="HCM3258"
gene
ORIGIN
Alignment Scores:
Pred. No.: 1,93e-214 Length: 1290
Score: 2055.00 Matches: 381
Percent Similarity: 90.97% Conservative: 2
Best Local Similarity: 90.50% Mismatches: 38
Query Match: 88.81% Indels: 0
DB: 10 Gaps: 0
US-10-602-219-12 (1-421) x AY408541 (1-1290)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAsnGlyLeuAlaArgThrPro 40
Db 61 CTCCTTTCTGGGACATCCCTTGGGCTAGAGCAGTGGCAATGGATTTGCAAGGAGCGCT 120
QY 41 ThrMetGlyTyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCATCTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
```

Db 181 GATTCTCGCATCAGTGAAGACTTTCATGGAGATGCGAGACTCATGTCTCNGAAGGC 240

QY 81 TTPLYAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100

Db 241 TGAAGGATGCGAGTTATGAGTACCTCTCGATTGATGACTGTGGATGGCTCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120

Db 301 GATTTCAGAGCGAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGATTCGNAGCTA 360

QY 121 AlaAsnTyrValHisSerIysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140

Db 361 NNNNNNNNGTTTCACAGCAAAGGACTGAAGCTAGGCAATTCACAGATCTCGAAATAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160

Db 421 ACCTGCGCAGGCTTCCTCGGAGTTTGGATACACTACGACATTCATGCCAGACCTTTCGT 480

QY 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180

Db 481 GACTGGGGAGTAGACTCTGTAATTTGATGGTTGTACTGTGACAGTTTGGANAATTG 540

QY 181 AlaAspGlyTyrIysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200

Db 541 CGACATGNN 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220

Db 601 NNN 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240

Db 661 CAGTACTGCAATCAGTGGCGAAATTTGTCTGAGTTGATGATCTCTGGAAAGATATAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnArgIleValAspValAlaGlyProGly 260

Db 721 AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAAATTTGTATGTTCTGCGACGAG 780

QY 261 GlyTrpAsnAspProAspMetIleValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280

Db 781 GGTTGGAATGACCCAGATATGTAGTATGTCGCACTTTGGCTCAGCTGGAAATCAGCAA 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300

Db 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTCTCTCTTTATCATGCTTAATGACCTC 900

QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320

Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCATCAAT 960

QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340

Db 961 CAGGACCCCTTGGCGAAGCAAGGATACCACTAGCTAGACAGNNAGACAACCTTGAAGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360

Db 1021 GAACGACCTCTCTCAGCTTAGCTGGCTGTAGCTGTATGATTAACCGCAGGAGATTGT 1080

QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380

Db 1081 GGACCTCGCTCTTATACCATCGCACTTCTTCCTCGGTAAAGGAGTGGCTGTAAATCCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400

Db 1141 GCCTGTCTTCATCACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCATGAATGGACT 1200

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuAsnThr 420

Db 1201 TCAAGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAAATACA 1260

QY 421 Met 421

Db 1261 ATG 1263

RESULT 6

AK054547

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AK054547 1376 bp mRNA linear HTC 03-APR-2004
Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN
full-length enriched library, clone:E330039P08
product:galactosidase, alpha, full insert sequence.

AK054547.1 GI:26344324
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861

4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6 (bases 1 to 1376)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, P., Imotani, K., Iehii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koyu, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

cdna library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.jp/>
URL: <http://fantom.gsc.riken.jp/>.

FEATURES

source

```
1..1376
/organism="Mus musculus"
/mol_type="mRNA"
/strains="C57BL/6J"
/db_xref="FANTOM DB:E330039P08"
/db_xref="taxon:10090"
/clone="E330039P08"
/sex="female"
/tissue_type="ovary"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="2 days pregnant adult"
21..1286
/note="unnamed protein product; galactosidase, alpha
(MGI:1347344, GB|NM_013463, evidence: BLASTN, 99%,
match=1373)
putative"
/codon_start=1
/protein_id="BAC35819.1"
/db_xref="GI:26344325"
/translation="MAMKLLSRDRLVCELALCPALVFWSLTGVRLDNLGIARTPTM
GWLHWRMCNLDCEEPDACSIOELFMQAEIYVSGWRDAGDYLCIDCWMAPE
DSKGLQADPFRPSGIEKHLAYVSKGLGIYADGNKTCAGPPGSGSYDIDAOT
FADKGVLLKEDGCHDSVSLNGYKYMALALNRGSIYVSCENPLYLPPFKPNY
TDIYCYCHNRNFDYDWSIESIKNLSVYQKEIVVEVAGPWNPDMLVIGNF
LSWQVQVMTALMAAPLLMSNDLRQISSQKALLQNDVIAINODPLGQGYCFR
KENHIEVWERPLSNLAWAVARNLQEIIGPCPYTTQISSLGRGLACNPGCIITOLLE
KVHLGFYEWTULTKTRVNSPGTVLFRLE"
```

CDS

```
1358..1363
/note="putative"
1376
/note="putative"
```

polyA_signal

polyA_site

ORIGIN

Alignment Scores:

```
Pred. No.: 1..4e-187 Length: 1376
Score: 1811.00 Matches: 327
Percent Similarity: 88.04% Conservative: 41
Best Local Similarity: 78.23% Mismatches: 50
Query Match: 78.26% Indels: 0
DB: 4 Gaps: 0
```

US-10-602-219-12 (1-421) x AK054547 (1-1376)

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 27 ATGAAGCTTTGAGCAGATACGCGGCTGGTCTGTGAGCTTGGCTTGTCCCTGGCT 86
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 87 TTAGTCTTCTGAGCATTCTTGGGGTCAGAGCATTGGCAATGGCTTGGCGGGACTCT 146
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 147 ACTATGGGCTGGCTGCATTGGGAAGCTTTTCATGTGCACCTTTGACTGCCAAGAAGGCT 206
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 207 GATGCTGCTAATGAGGAGCACTGTTTCATGAGTGGCAGAGCTCATGGTCTCTGATGGC 266
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 267 TGGCGGATGAGGTTATCACTCTCTCATAGTAGCTGTTGGATGGCTCCCGAGAGG 326
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 327 GATTCAAGGGGAGGCTTCAGGCAGATCCCAACGCTTTCTTAGTGGGATCAACACCTC 386
```

```
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 387 GCAAAATACGTCCACAGCAAGAGATTGAAGCTAGGATTATGACAGATTGTTGGGAATAAA 446
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 447 ACCTGTGCAGGTTTCCCGGGAGTTTGGATCCCTATGACATTGATGCCAGACATTGCT 506
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 507 GACTGGGCGTAGATCTGCTAAATTTGATGTTGTCTACTGTGACAGTGTAGTATCCTTGT 566
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 567 GAGAATGGTTATAAGTACATACGCTTGGCTTGAACAGGACAGCGGCGAAGCATTTGTAT 626
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 627 TCCTGTGAGTGGCCACTTTATTTGAGACCTTTTCATAAGCCCAATATACAGATATCCNA 686
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 687 TATTACTGCATCATTTGGAGAAATTTTCATGATGTTTATGATCTCTGGGAAGCATAAAG 746
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 747 AATATCTTGTCTTGACAGAGTGGTTTACCAGAAAGAGAGATTGTTGAAGTCGCTGACCGC 806
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 807 AGCTGGAATACCCAGACATGTTAGTATGCGCAACTTTGGCTCAGTTGGGACGACGAG 866
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 867 GTGACACAGATGGCCCTCTGGGCTATCATGGCCGCTCCCTACTCATGTGCCAAGCATCTG 926
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 927 CGACAAATCAGCTCTCAAGCCAAAGCTCTGCTCAGATAAGAGTGTAAATGCCCATCAAC 986
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 987 CAAGACCCCTTGGGCAAGCAGCGCTACTGTTTCAGAAAGAGAAACCATATTGAGGTTGG 1046
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1047 GAACGGCCACTCTCCAACTAGCTGGGCTGTGGCTGTGAGAAACCTCGACGAGATTGGT 1106
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1107 GGACCTTGTCTTATACCATCCAGATTTCCTCCCTGGGTAGAGGACTAGCTGCAATCCT 1166
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1167 GGCTGCATCATTTACTCAGCTTCTCCCCGAGAAAGTACACCTAGGCTTCTATGAATGGACT 1226
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeuGlu 418
Db 1227 TTGACCTTAAAACTCGAGTAACCCCTCAGGCACTGTTTGTTCGGTTAGAA 1280
```

RESULT 7

```
AK040081 2962 bp mRNA linear HTC 03-APR-2004
LOCUS Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched
DEFINITION library, clone:A430057F16 product:galactosidase, alpha, full insert
sequence.
ACCESSION AK040081
VERSION AK040081.1 GI:26333580
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```



```

Db      607  TCCCTGAGTGGCCACTTATTTAGAGACCTTTCATAAGCCCAATTATACAGATATCCAA 666
Qy      221  GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db      667  TATTACTGCAATCATTTGGAGAAATTTTGTATGATGTTTATGATCTCTCGGAAGCATAAAG 726
Qy      241  SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db      727  AATACTTGTCTTGGACAGTGGTTTACAGAGAGAGATTGTTGAAGTCGTGACCAAGGC 786
Qy      261  GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db      787  AGCTGGAATGACCCAGACATGTTAGTGTGCGCAACTTTTGGCCTCAGTTGGGACGACG 846
Qy      281  ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db      847  GTGACACAGATGGCCCTCTGGGCTATCATGGCGCTCCCTTACTCATGTCCCAACCATCTG 906
Qy      301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      907  CGCAAAATCAGCTCTCAAGCCAAAGCTCTGCTTTCAGATAAGGATGTAATTGCCATCAAC 966
Qy      321  GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db      967  CAAGACCCCTTGGCAAGCAGCGCTACTGTTTCAGAAAGGAAACCAACATTCAGGTTGG 1026
Qy      341  GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnLysGly 360
Db      1027  GAACGGCCACTCTCAACCTACCCCTGGGCTGTGGCTGTGAGAAACCTCGCAGGAGATTGGT 1086
Qy      361  GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1087  GGACCTTGTCTTATACCATCAGATTCTTCCCTGGGTAGAGGACTAGCTGCATCTCT 1146
Qy      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db      1147  GGCTGCATCTTACTCAGCTTCTCCCGAGAAAGTACACTAGGCTTCTTATGAATGGACT 1206
Qy      401  SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGlu 418
Db      1207  TTGACCTTAAAACTCGAGTAAACCCCTCAGGCACTGTTTGTTCGGTGTAGAA 1260

RESULT 8
AL554978
LOCUS   AL554978 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
DEFINITION
AL554978 cDNA clone CSODK009YB12 5-PRIME, mRNA sequence.
ACCESSION
AL554978.3 GI:45859723
KEYWORDS
EST.
SOURCE  Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 1086)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:31276787.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: segrefgenoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CSODK009B06QP1&c=10506.r.

```

FEATURES

```

source
1..1086
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODK009YB12"
/cell_type="HELA"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Alignment Scores:
Pred. No.: 5,33e-184 Length: 1086
Score: 1777.00 Matches: 342
Percent Similarity: 96.65% Conservatives: 4
Best Local Similarity: 95.53% Mismatches: 10
Query Match: 76.79% Indels: 7
DB: 1 Gaps: 1
US-10-602-219-12 (1-421) x AL554978 (1-1086)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 6 ATGCAGCTGAGGAAACCCAGACTACATCTGGCTGGCGCTTGC-CTTCGCTTCCTGGCC 64
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 65 CTCGTTTCTCTGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 124
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 125 ACCATGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 184
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 185 GATTCCTGCATCAGTGAAGCTCTTCATGCAGATGGCAGAGCTCATGTCTCAGAAAGC 244
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 245 TCGAAGGATGAGGTTATGAGTACCTCTGCATTTGATGATGATGATGATGATGATGAT 304
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 305 GATTCAGAAAGGAGACATTCAGGACAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 364
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 365 GCTAATTTATTTTCAGCAAGAGCTGAAGCTAGGATTTATGCAGATGTTTGGAAATAA 424
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 425 ACCTCGCAGGCTTCCCTGGGAGTTTGGTACTACTACGACATTCATGATGATGATGATGAT 484
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 485 GACTGGGAGTAGATCTCTAAATTTTATGATGATGATGATGATGATGATGATGATGAT 544
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 545 GCAGATGTTTATAGCACATGTCTTGGCCCTGATAGACTGGCAGAGACATTTGTGTAC 604
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 605 TCCTGTGAGTGGCTCTTTTATATGTTGGCCCTTTTCAAAAGCCCAATATATACAGAAATCCGA 664
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 665 CAGTACTGCAATCACTGGCGAAATTTTCTGTCATTTGATGATTTCTCTGAAAGATATAAG 724

```

```

QY 241 SerIleLeuAspTrpThrSerPheAenGlnGluArqIleValAspValAlaGlyProGly 260
Db 725 AGTATCTTGACCTGGACATCTTTTAACAGAGGAGAAATTTGTGATGTTCTGGACGAGG 784

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 785 GGTGGGAATGACCAAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 844

QY 281 ValThrGlnMetAlaLeuTrpAlaIle-MetAlaAlaProLeuPheMetSerAsnAspLe 300
Db 845 GTAACACTCAGATGCCCTCTGGGCTATTCATGGCTCTCTCTTATTCAATGCTTAATGACCT 904

QY 300 uArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAs 320
Db 905 CCGACATCATGCCCTCAGCCAAAGCTCTCTTCAGGATAGGAGCTWATTTGCCATCAA 964

QY 320 nGlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTr 340
Db 965 TCAGGACCCVTK-GGCAAGCAAGGTACAGCTTAGAMAGGGGRCACAM-TTTGWAAGTGTG 1022

QY 340 pGluArgProLeuSerGlyLeuAlaTrpAlaMetIleAsnArgGln 357
Db 1023 GGA-MGA---CTCTCTCAGGTAGCTGGCTGTAGY-ATGATAAACCCGAGG 1069

RESULT 9
BX354096 1014 bp mRNA linear EST 23-APR-2004
LOCUS BX354096 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DC014YH04 5-PRIME, mRNA sequence.
ACCESSION BX354096
VERSION BX354096.2 GI:46550107
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1014)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 5, 2003 this sequence version replaced gi:30379845.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.x
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DC014DD02QP1&c=10506.r.

FEATURES
Location/Qualifiers
1..1014
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC014YH04"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Alignment Scores: 1.8e-175 Length: 1014
Pred. No.: 1699.00 Matches: 317
Score: 1699.00
Percent Similarity: 97.54% Conservative: 0
Best Local Similarity: 97.54% Mismatches: 7

QY 5 AsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAlaLeuValSerTrp 24
Db 31 AACCCAGAACTACATCTGGGCTGCGGCTTGCSSTT-CGCTTCCTGGCCCTCGTTTCCTGG 89

QY 25 AspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrp 44
Db 90 GACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCCCTACCATGGGCTGG 149

QY 45 LeuHisTrpGluArqPheMetCysAsnLeuAspCysGlnGluGluProAspSerCysIle 64
Db 150 CTGCACTGGGAGCGCTTCATGTCAACCTTGACTGCCAGGAGAGCCAGATTCTCGCATC 209

QY 65 SerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLysAspAla 84
Db 210 AGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGCTGGAGAGTGA 269

QY 85 GlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSerGluGly 104
Db 270 GGTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCCAAAGAGATTCAAGAGGC 329

QY 105 ArgIleuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeuAlaAsnTrpVal 124
Db 330 AGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGGATTCGCCAGCTAGCTAATATTGTT 389

QY 125 HisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLysThrCysAlaGly 144
Db 390 CACACCAAGAGACTGAGACTAGGATTTATGCAGATGTTGGAAATATAAACCTGCCAGGC 449

QY 145 PheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAlaAspTrpGlyVal 164
Db 450 TTCCCTGGGAGTTTGGTACTACGACATTGATGCCAGACTTTCCTGCTGACTGGGGAGTA 509

QY 165 AspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeuAlaAspGly-Ty 184
Db 510 GATCTGTAAAATTTGATGGTGTGTACTGTGACAGTTTGGAAAATTTGGCAGATGGTATA 569

QY 184 rLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrpSerCysGluTr 204
Db 570 TAAGCACATGCTCTTGGCCCTGATAGACTGGCAGAGCATTTGTGTACTCTCTGTGAGTG 629

QY 204 pProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArgGlnTrpCysAs 224
Db 630 GCCTCTTTATATGTGGCCCTTTCAAAGGCCCAATTTATACAGAAATCCGACAGTACTGCAA 689

QY 224 nHisTrpArgPheAlaAspIleAspAspSerTrpLysSerIleLysSerIleLeuAs 244
Db 690 TCACCTGGCGAAATTTTGTGACATTGATGATTCCTGGAAAAGATATAAAGAGTATCTTGA 749

QY 244 pTrpThrSerPheAenGlnGluArqIleValAspValAlaGlyProGlyGlyTrpAsnAs 264
Db 750 CTGGACATCTTTTAACAGAGAGAGAAATTTGTGATGTTCTGGACAGCGGGGTGGAAATGA 809

QY 264 pProAspMetLeuValIleGlyAsnPheGlyLysSerTrpAsnGlnGlnValThrGlnMe 284
Db 810 CCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAAATCAGCAAGTAACCTCAGAT 869

QY 284 tAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSe 304
Db 870 GGCCCTCTGGGGCTATCATGGCTGCTCTTTTATTCAATGCTTAATGCTCCGACACATCAG 929

QY 304 rProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsnGlnAspProle 324
Db 930 CCCYCAAGCCAAAGCTCTCTTCAGRTAAGRCGTAAATTTGCCAC-AWCAGAGCCCTK 988

QY 324 uGlyLysGlnGly 328
Db 989 -GGCMAGCAAGGK 1000

```

```

Query Match: 73.42% Indels: 4
DB: 5 Gaps: 0
US-10-602-219-12 (1-421) x BX354096 (1-1014)

QY 5 AsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAlaLeuValSerTrp 24
Db 31 AACCCAGAACTACATCTGGGCTGCGGCTTGCSSTT-CGCTTCCTGGCCCTCGTTTCCTGG 89

QY 25 AspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrp 44
Db 90 GACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCCCTACCATGGGCTGG 149

QY 45 LeuHisTrpGluArqPheMetCysAsnLeuAspCysGlnGluGluProAspSerCysIle 64
Db 150 CTGCACTGGGAGCGCTTCATGTCAACCTTGACTGCCAGGAGAGCCAGATTCTCGCATC 209

QY 65 SerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLysAspAla 84
Db 210 AGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGCTGGAGAGTGA 269

QY 85 GlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSerGluGly 104
Db 270 GGTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCCAAAGAGATTCAAGAGGC 329

QY 105 ArgIleuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeuAlaAsnTrpVal 124
Db 330 AGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGGATTCGCCAGCTAGCTAATATTGTT 389

QY 125 HisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLysThrCysAlaGly 144
Db 390 CACACCAAGAGACTGAGACTAGGATTTATGCAGATGTTGGAAATATAAACCTGCCAGGC 449

QY 145 PheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAlaAspTrpGlyVal 164
Db 450 TTCCCTGGGAGTTTGGTACTACGACATTGATGCCAGACTTTCCTGCTGACTGGGGAGTA 509

QY 165 AspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeuAlaAspGly-Ty 184
Db 510 GATCTGTAAAATTTGATGGTGTGTACTGTGACAGTTTGGAAAATTTGGCAGATGGTATA 569

QY 184 rLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrpSerCysGluTr 204
Db 570 TAAGCACATGCTCTTGGCCCTGATAGACTGGCAGAGCATTTGTGTACTCTCTGTGAGTG 629

QY 204 pProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArgGlnTrpCysAs 224
Db 630 GCCTCTTTATATGTGGCCCTTTCAAAGGCCCAATTTATACAGAAATCCGACAGTACTGCAA 689

QY 224 nHisTrpArgPheAlaAspIleAspAspSerTrpLysSerIleLysSerIleLeuAs 244
Db 690 TCACCTGGCGAAATTTTGTGACATTGATGATTCCTGGAAAAGATATAAAGAGTATCTTGA 749

QY 244 pTrpThrSerPheAenGlnGluArqIleValAspValAlaGlyProGlyGlyTrpAsnAs 264
Db 750 CTGGACATCTTTTAACAGAGAGAGAAATTTGTGATGTTCTGGACAGCGGGGTGGAAATGA 809

QY 264 pProAspMetLeuValIleGlyAsnPheGlyLysSerTrpAsnGlnGlnValThrGlnMe 284
Db 810 CCCAGATATGTTAGTATGGCACTTTGGCCTCAGCTGGAAATCAGCAAGTAACCTCAGAT 869

QY 284 tAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSe 304
Db 870 GGCCCTCTGGGGCTATCATGGCTGCTCTTTTATTCAATGCTTAATGCTCCGACACATCAG 929

QY 304 rProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsnGlnAspProle 324
Db 930 CCCYCAAGCCAAAGCTCTCTTCAGRTAAGRCGTAAATTTGCCAC-AWCAGAGCCCTK 988

QY 324 uGlyLysGlnGly 328
Db 989 -GGCMAGCAAGGK 1000

```

RESULT 10

AL552630 999 bp mRNA linear EST 30-MAR-2004
 LOCUS AL552630 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
 DEFINITION clone CS01067YJ01 5-PRIME, mRNA sequence.

ACCESSION AL552630
 VERSION AL552630.3 GI:45857411

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.

REFERENCE

1 (bases 1 to 999)

AUTHORS

Li.W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

On Feb 15, 2001 this sequence version replaced gi:31274445.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 2 rue Gaston Cremieux, Cp 5706 - 91057 EVRY cedex - FRANCE
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with NotI and cloned
 into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 10506.r

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna/s=CS0D1067CE01QPI&c=10506.r.

FEATURES

source

1..999

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0D1067YJ01"

/tissue_type="PLACENTA COT 25-NORMALIZED"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 1,17e-174 Length: 999
 Score: 1691.50 Matches: 323
 Percent Similarity: 98.48% Conservative: 0
 Best Local Similarity: 98.48% Mismatches: 5
 Query Match: 73.10% Indels: 5
 DB: 1 Gaps: 0

US-10-602-219-12 (1-421) x AL552630 (1-999)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 7 ATGCAGCTGAGGACCCAGCACTACATCTGGCGCTGGCGCTGGCGTTCGCTTCCTGCC 65
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 66 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAGGACGCCT 125
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 126 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGATGCCAGGAAGGCCA 185
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 186 GATTCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 245
 QY 81 TrpLysAspAlaGlyTrpGluTrpLeuGlyCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 246 TGGAGGATGCAAGGTATGAGTACCTCTGCAATTGATGACTGTGGATTGGCTCCCAAGA 305

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 306 GATTCAAGAGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGATTTCGCAGCTA 365
 QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
 Db 366 GCTAATTATGTTTACAGCAAGGAGCTGAAGCTAGGGATTTATGAGATGTTGGAAATAAA 425
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
 Db 426 ACCTGGCAGGCTTCCTCGGAGTTTGGATCTACGACATTTGATGCCAGACCTTTGCT 485
 QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 Db 486 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTTACTGTGACAGTTTGGAATTTG 545
 QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
 Db 546 GCAGATGGTTATAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 605
 QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
 Db 606 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 665
 QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 666 CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTGATGATTCCTGGAAGATATAAG 725
 QY 241 SerIleLeuAspTrpTrpSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 726 AGTATCTTGGACTGGACATCTTTTAACCAAGAGAGAAATTTGTATGTTGCTGGACACGG 785
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 786 GGTGGATGATGACCAAGATATGTTAGTATGGCAACTTTGGCCTCAGCTGGATCAGCAA 845
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 846 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTCTCTCTCTCTCTCTCTCT 904
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 905 CGACACATCAGCCCTCAGCCCAAGCT-CTCCTTCAGGATAAGGACGTAATTTGCAWC-AAT 962
 QY 321 GlnAspProLeuGlyLysGlnGly 328
 Db 963 CAGGACCCCTTG-GGCAAGCAAGGK 985

RESULT 11

BM564282

LOCUS

BM564282

DEFINITION

5', mRNA sequence.

ACCESSION

BM564282

VERSION

BM564282.1 GI:18811955

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.

REFERENCE

1 (bases 1 to 1067)

AUTHORS

NIH-MSC http://mgc.nci.nih.gov/.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be

BQ062192 1005 bp mRNA linear EST 02-APR-2000
 AGENCOURT_6829871 NIH_MGC_99 Homo sapiens cdna clone IMAGE:5923640
 5', mRNA sequence.
 BQ062192
 BQ062192.1 GI:19888802
 EST.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 1 (bases 1 to 1005)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Lou Staudt
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM2092 row: h column: 09
 High quality sequence stop: 676.
 Location/Qualifiers
 1..1005
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5923640"
 /isue_type="lymphoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_99"
 /notes="Organ: lymph; Vector: pOT77; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:"

GCACAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 3,89e-171 Length: 1005
Score: 1659.50 Matches: 312
Percent Similarity: 96.63% Conservatives: 3
Best Local Similarity: 95.71% Mismatches: 5
Query Match: 71.72% Indels: 6
DB: 3 Gaps: 1

US-10-602-219-12 (1-421) x BQ062192 (1-1005)

QY 28 GlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrpLeuHisTrp 47
DB 1 GGGGCTAGAGCACTGGCAATGGATTGGCAGAGCGCTTACCATGGGCTGGCTGCACTGG 60
QY 48 GluArgPheMetCysAsnLeuAspCysGlnGluProAspSerCysIleSerGluLys 67
DB 61 GAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCAGATTCTGTCATCAGTGAGAG 120
QY 68 LeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLysAlaGlyTrpGlu 87
DB 121 CTCCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAGGCTGGAAGATGCAGTTATGAG 180
QY 88 TyrLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGln 107
DB 181 TACCTCTGATGATGACTGTTGGATGGCTGCCCAAGAGATTCAAGAGGCAGACTTCAG 240
QY 108 AlaAspProGlnArgPheProHisGlyLeuArgGlnLeuAlaAsnTyrValHisSerLys 127
DB 241 CGAGACCTTCAGCGCTTCCTCATGGGATTCGCCAGTAGCTAAATATGTTACAGCAAAA 300
QY 128 GlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLysThrCysAlaGlyPheProGly 147
DB 301 GGACTGAAGCTAGGATTTATGCAGATGTTGGAATTAACCTGGCAGGCTTCCTGGG 360
QY 148 SerPheGlyTyrTyrAspIleAspAlaGlnThrPheAlaAspTrpGlyValAspLeuLeu 167
DB 361 AGTTTGGATATACGACATTCATGCCACACCTTTGCTGACTGGGGAGTAGATCTGCTA 420
QY 168 LysPheAspGlyCysTyrCysAspSerLeuGluAsnLeuAlaAspGlyTyrIleHisMet 187
DB 421 AAAATTGATGGTTGTTACTGTGACACAGTTTGGAAAATTTGGCAGATGGTTATAAGCACATG 480
QY 188 SerLeuAlaLeuAsnArgThrGlyArgSerIleValTyrSerCysGluTrpProLeuTyr 207
DB 481 TCCTTGGCCCTGANTAGGACTGGCAGAGCATGTGTTCTCTGTGAGTGGCTCTTTAT 540
QY 208 MetTrpProPheGlnLysProAsnTyrThrGluIleArgGlnTyrCysAsnHisTrpArg 227
DB 541 ATGTGGCCCTTTCAAAAGCCCAATATATACAGAAATCCGACAGTACTGCAATCACTGGCGA 600
QY 228 AsnPheAlaAspIleAspAspSerTrpLysSerIleLysSerIleLeuAspTrpThrSer 247
DB 601 AATTTTGTCTGACATTCATGATTCCTCGAAAAGATATAAAGAGATATCTTGAGCTGGACATCT 660
QY 248 PheAsnGlnGluArgIleValAspValAlaGlyProGlyGlyTrpAsnAspProAspMet 267
DB 661 TTTTACCAGAGAGAAATTTGTTGATGTTGCTGCACAGGGGGTTGGAATGACCCAGATATG 720
QY 268 LeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGlnValThrGlnMetAlaLeuTrp 287
DB 721 TTATGATTTGGCAACTTTTGGCTCAGCTGCAATCAGCAAGTAACCTCAGATGGCCCTCTGG 780
QY 288 AlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSerProGlnAla 307
DB 781 GCTATCATGGCTGCTCTTTTATCATGTTATATGACCTCCGACATCAGCCCTCAAGCC 840

QY 308 LysAlaLeuLeuGlnAspLysAspValIleAla-IleAsnGlnAspProLeu-GlyLysG 327
DB 841 AAAGCTCTCTTCAGGATAAAGAGCTTATCCCATCAATCAGGACCCCTTTGGGGCAAGC 900
QY 327 In-GlyTyrGlnLeu-ArgGlnGly-AspAsnPheGlu---ValTrpGluArgProLeuS 345
DB 901 AAAGTTTACCAGCTTTTAAACAGGGAACCACTTTTGAAGGGTGGGAACGACCCCTCTCT 960
QY 345 erGlyLeu 347
DB 961 CAGGGCTA 968
RESULT 13
LOCUS CO645623
DEFINITION ILLUMIGEN MCQ 24642 Katze MMLV Macaca mulatta cDNA clone
IBIUM_23923 5' similar to Bases 5 to 920 highly similar to human
GLA (Hs.69089), mRNA sequence.
ACCESSION CO645623
VERSION CO645623.1 GI:50567117
KEYWORDS EST.
SOURCE Macaca mulatta (rhesus monkey)
ORGANISM Macaca mulatta
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Cercopithecoidea; Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 922)
AUTHORS Magnes, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., Agy, M.B.,
Prohl, S.C., Fitzgibbon, M., Scherer, C.A., Minet, D.G., Katze, M.G. and
Iadonato, S.P.
TITLE Analysis of the Macaca mulatta transcriptome and the sequence
divergence between Macaca and human
Genome Biol. 6 (7), R60 (2005)
JOURNAL 15998449
PUBMED Contact: C. Magnes
ILLUMIGEN Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagnes@illumigen.com
Sequenced on 2004.03.02. 796 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
ILLUMIGEN Biosciences Inc. For further information, see
http://www.macaque.org
PCR Primers
FORWARD: CCCTCACTAAAGGGAACAAA
BACKWARD: CACTATAGGCGAATTGGGTA
Insert Length: 922 Std Error: 0.00
Plate: CL000111 row: E column: 08
Seq primer: CCCTCACTAAAGGGAACAAA
POLYA=No.
FEATURES Location/Qualifiers
1..922
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="IBIUM:23923"
/sex="female"
/dev_stage="adult"
/lab_host="Electromax DH10B"
/clone_lib="Katze MMLV"
/note="Organ: liver; Vector: pDONR 222; Site 1: BsrG I;
Site 2: BsrG I; Created from Cloneminer cDNA Library
Construction kit (catalog #18249-029)"
ORIGIN
Alignment Scores:
Pred. No.: 4,86e-170 Length: 922
Score: 1649.00 Matches: 295
Percent Similarity: 98.03% Conservatives: 3
Best Local Similarity: 97.04% Mismatches: 6

Query Match:	71.26%	Indels:	0
DB:	7	Gaps:	0
US-10-602-219-12 (1-421) x C0645623 (1-922)			
Qy	1	MetGlnLeuArgAnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
Db	11	ATGCAGCTGAGAAACCCAGAAATGATCATCTGGGCTGCGCTTGGCGCTTCTCTCTTCGTGCC	70
Qy	21	LeuValSerTrpAspIleProGluValaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
Db	71	CTCGTTTCTGGGACATCCCTGGGGCCAGAGCATGGACAATGGATTGGCCAGGACGGCT	130
Qy	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
Db	131	ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCA	190
Qy	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
Db	191	GATTCTCGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTTCAGATGGC	250
Qy	81	TrpLysAspAlaGlyTyrgluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg	100
Db	251	TGGAAGGATGCAAGTTATGATGACTCTGCATTTGATGACTGTTGGATGGCTCCCCAAGA	310
Qy	101	AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
Db	311	GATTTAGAAGGCAGACTTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA	370
Qy	121	AlaSerTrpValHisSerLysGlyLeuLysLeuGlyIleTyrgluAlaAspValGlyAsnLys	140
Db	371	GCTAATTATGTTTACAGCAAAAGGACTGAAGCTAGGGATTTATGTCAGATGTTTGGAAATAA	430
Qy	141	ThrCysAlaGlyPheProGlySerPheGlyTyrgluTrpAspIleAspAlaGlnThrPheAla	160
Db	431	ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACTATGACATGATGGTCCAGACCTTTGTCT	490
Qy	161	AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu	180
Db	491	GACTGGGAGTAGACTCTGCTGAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAAGTTG	550
Qy	181	AlaAspGlyTyrgluHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyrglu	200
Db	551	GCAGATGTTTATAGACATGCTCTTGGCCCTTGAATAGGACTGGCAGAGCAATTCGTGAC	610
Qy	201	SerCysGluTrpProLeuTyrgluTrpProPheGlnLysProAsnTyrgluTrpLeuArg	220
Db	611	TCCTGTGAGTGCCCTCTTTATATGTGGCCATTTCAAAAGCCCCAATTACACAGAAATCCGA	670
Qy	221	GlnTyrgluAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys	240
Db	671	CAGTACTGCATCACTGCGGAAATTTTGTGTCATGATGATTCCTGGNAAAGTATTAAG	730
Qy	241	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValaGlyProGly	260
Db	731	AGTATCTTGGACTGCACATCTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACCGAGG	790
Qy	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln	280
Db	791	GGTTGGAAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCCTCAGCTGGGAATCAGCA	850
Qy	281	ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu	300
Db	851	GTAACCTCAGATGGGCGCTTCGCTATCATGGCTGCTCCCTTTATTCATGCTTAATGACCTC	910
Qy	301	ArgHisIleSer	304
Db	911	CGACACATCACC	922
RESULT 14			
AL577581/c			
LOCUS			
DEFINITION AL577581 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens			
linear EST 07-APR-200			

ACCESSION	CDNA clone CS0DK009YE12 3-PRIME, mRNA sequence.
VERSION	AL577581
KEYWORDS	AL577581.3 GI:46256603
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 1026)
TITLE	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
JOURNAL	Full-length cDNA libraries and normalization
COMMENT	Unpublished (2001) On Feb 16, 2001 this sequence version replaced gi:31315840. Contact: Genoscope Genoscope - Centre National de Sequencage 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 10506.r For more information about this cluster, see http://www.genoscope.cns.fr/cdna?s=CS0DK009BC06GNP1&c=10506.r. Location/Qualifiers 1. .1026 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0DK009YE12" /cell_type="HELA CELLS COT 25-NORMALIZED" /cell_line="HELA" /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED" /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
FEATURES	source

[illegible]

RESULT 14	AL577581/c	AL577581	1026 bp	mRNA	linear	EST 07-APR-2004
		AL577581				
		DEFINITION	Homo sapiens HELA CELLS COT 25-NORMALIZED	Homo sapiens		


```
Db 657 AGCATGTGTACTCTCTGTGAGTGGCCCTTTATATGTGGCCCTTTCAAAAGCCCAATTAT 598
Qy 217 ThrGluileArgGlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspSerTrp 236
Db 597 ACAGAAATCCGACAGTACTGCAATCACTGGCGAATTTTGTGACATTTGATGATTCCTGG 538
Qy 237 LysSerIleLysSerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspVal 256
Db 537 AAAAGTATAAAGAGTATCTTGGACTGGACATCTTTTAAACGAGGAGAGAAATTTGTGATGT 478
Qy 257 AlaGlyProGlyGlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSer 276
Db 477 GCTGACCAGAGGGGTGGATGACCCAGATATGTTAGTGAATGGCACTTTGGCCCTCAGC 418
Qy 277 TrpAsnGlnGlnValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMet 296
Db 417 TGGATATCAGCAAGTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTATG 358
Qy 297 SerAsnAspLeuArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspVal 316
Db 357 TCTAATGACCTCCGACATCAGCCCTCAAGGCCAAAGCTCTCTCTCAGGATAAGGACGTA 298
Qy 317 IleAlaIleAsnGlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsn 336
Db 297 ATTGCCATCAATCAGGACCCCTTGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAAC 238
Qy 337 PheGluValTrpGluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArg 356
Db 237 TTTGAGTGTGGGAACGACCTCTCTCAGCTTAGCTGGCTGTAGCTATGATATAACCGG 178
Qy 357 GlnGluileGlyGlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyVal 376
Db 177 CAGGAGATTGGTGGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGTAAAGGAGTG 118
Qy 377 AlaCysAsnProAlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPhe 396
Db 117 GCCTGTAATCTCCCTGCTTTCATCACAGCTCCCTCCMTRTGAAAGGAAGCTAGGGTTC 58
Qy 397 TyrGluTrpThrSerArgLeuArgSerHisIleAsnProThr 410
Db 57 TATGATGGACTTCAAGTTAGAGTAGTAACATAATVCRACA 16
```

RESULT 15

```
AY408542
LOCUS      1296 bp      DNA      linear      GSS 15-DEC-2003
DEFINITION Mus musculus CIA gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
```

```
ACCESSION  AY408542
VERSION     AY408542.1  GI:39764513
```

```
KEYWORDS   GSS.
```

```
SOURCE     Mus musculus (house mouse)
```

ORGANISM

```
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
```

```
1 (bases 1 to 1296)
```

```
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
```

```
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
```

```
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
```

```
Adams, M.D. and Cargill, M.
```

```
Adams, M.D. and Cargill, M.
```

```
Direct Submission
```

```
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
```

```
Rockville, MD 20850, USA
```

```
This sequence was made by sequencing genomic exons and ordering
```

TITLE

```
Inferring nonneutral evolution from human-chimp-mouse orthologous
```

```
gene trios
```

```
Science 302 (5652), 1960-1963 (2003)
```

```
14671302
```

```
2 (bases 1 to 1296)
```

```
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
```

```
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
```

```
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
```

```
Adams, M.D. and Cargill, M.
```

```
Direct Submission
```

```
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
```

```
Rockville, MD 20850, USA
```

```
This sequence was made by sequencing genomic exons and ordering
```

COMMENT

```
FEATURES             them based on alignment.
     source            Location/Qualifiers
     1..1296
     /organism="Mus musculus"
     /mol_type="genomic DNA"
     /db_xref="taxon:10090"
     <1..>1296
     /gene="CIA"
     /locus_tag="HMC3258"
```

ORIGIN

```
Alignment Scores:
Pred. No.:      6,04e-168      Length:      1296
Score:          1632.00        Matches:      301
Percent Similarity: 81.10%     Conservative: 38
Best Local Similarity: 72.01%   Mismatches:   79
Query Match:     70.53%        Indels:        0
DB:              10           Gaps:         0
```

```
US-10-602-219-12 (1-421) x AY408542 (1-1296)
```

```
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
```

```
Db 1 ATGAAGCTTTTGACGACAGATACCGGCTGTCTGTGAGCTTTGGCTTTGCTCCCTGGCT 60
```

```
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
```

```
Db 61 TTAGTTTCTGGAGCATCTTGGGGTCAGAGCAATGGACAATGGCTTGGCGCGGACTCCT 120
```

```
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluLeuPro 60
```

```
Db 121 ACTATGGCTGGCTGGCATTTGGAAACGTTTCATGTGCAACCTTGACTGCCAAGAAGAGCT 180
```

```
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
```

```
Db 181 GATSCCTGCAATAGTGAGCAACTGTTCATGCAGATGGCAGAGCTCATGCTCTCTGATGGC 240
```

```
Qy 81 TrpLysAspAlaGlyTyrGluTyrIleuCysIleAspAspCysTrpMetAlaProGlnArg 100
```

```
Db 241 TGGCGGATGCAGGTTATGACTATCTCTGCATAGTACTGTTGGATGGCTCCCGAGAGG 300
```

```
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
```

```
Db 301 GATTCAAGGCGCAGCTTTCAGGACAGATCCCAACGCTTTCTAGTGGGATCAACACCTC 360
```

```
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
```

```
Db 361 GCAATTTACGTCACAGCAAGGATTTAGAGTAGGATTTATGCAGATGTTGGGATATAA 420
```

```
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
```

```
Db 421 ACCTGTGCAGGTTTCCCGGAGTTTGGATCCTATGACATTTGATGCCGACACATTTGCT 480
```

```
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
```

```
Db 481 GACTGGGCGGTAGATCTGCTAAATTTGATGGTTGTCTACTGTGACAGTAGTAGTCTG 540
```

```
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
```

```
Db 541 GAGATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600
```

```
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
```

```
Db 601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660
```

```
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
```

```
Db 661 TATTACTGCAATCATTTGGAGAAATTTTTCATGATGTTTATGATCTCTGGGAAGCATAAAG 720
```

```
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
```

```
Db 721 AATATCTTGTCTGGACAGAGTGTATTACCAAGGAGGAGTGTGTTGAAGTCTCTGGACCGC 780
```

```
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db |||||
781 AGCTGGAATGACCCAGACATGTTAGTGATCGCAACTTTGGCCTCAGTTGGGACCCAGCAG 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db |||||
841 GTGACACAGATGGCCCTCTGGGCTATCATGGCGCTCCCTACTCATGTCCAACGATCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db |||||
901 CGACAATCAGCTCTCAAGCCAAAGCTCTGCTTCAGATAAGGATGTAATGCCATCAAC 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlnGlyAspAsnPheGluValTrp 340
Db |||||
961 CAAGACCCCTTGGGCAAGCAGGGCTACTGTTTCAGAAAGGAAACCCACATTGAGGTTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db |||||
1021 GAACGGCCACTCTCCAACCTAGCCTGGGCTGTGGCTGTGAGAAACCTGCAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db |||||
1081 GGACCTTGCTCTTATACCATCCAGATTTCTCCCTGGGTAGAGGACTAGCCTGCAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db |||||
1141 GGCTGCATCATTTACTCAGCTTCTCCCGAGAAAGTACACTAGGCTTCTATGATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGlu 418
Db |||||
1201 TTGACCTTAAAAAATCGAGTAAACCCCTCAGGCACCTGTTTGTTCGGTTAGAA 1254
```

Search completed: December 26, 2005, 13:48:40
Job time : 4103 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 26, 2005, 08:34:41 ; Search time 666 Seconds

(without alignments)
4212.966 Million cell updates/sec

Title: US-10-602-219-12

Perfect score: 2314

Sequence: 1 MQLRNPELHGCALALRFLA.....RLRSHINPTGTVLQLENTM 421

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10602219/runat_23122005_151141_17129/app_query.fasta_1.583
-DB=N_Geneseq -QFWT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORES=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10602219 @CGN 1 1 1096 @runat_23122005_151141_17129 -NCPU=6 -ICPU=3
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq 21.1*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2314	100.0	1266	6	AAd45223 Human rGA
2	2314	100.0	1266	10	Ad484750 Human alp
3	2314	100.0	1266	12	Adj88276 Human WT
4	2314	100.0	1266	12	Adm48680 Human wil

5	2314	100.0	1266	13	ADU66915	Adu66915 Human alp
6	2314	100.0	1266	14	AEA27444	Aea27444 Human alp
7	2314	100.0	1278	6	AAd45221	Aad45221 Human rGA
8	2314	100.0	1278	10	ADd84746	Ad84746 Human rGA
9	2314	100.0	1278	12	ADJ88272	Adj88272 Human rGA
10	2314	100.0	1278	12	ADm48676	Adm48676 Human wil
11	2314	100.0	1278	13	ADU66911	Adu66911 Human alp
12	2314	100.0	1278	14	AEA27440	Aea27440 Human alp
13	2314	100.0	1284	6	AAd45224	Aad45224 Human rGA
14	2314	100.0	1284	10	ADd84752	Ad84752 Human alp
15	2314	100.0	1284	12	ADJ88278	Adj88278 Human WT
16	2314	100.0	1284	12	ADm48682	Adm48682 Human wil
17	2314	100.0	1284	13	ADU66917	Adu66917 Human alp
18	2314	100.0	1284	14	AEA27446	Aea27446 Human alp
19	2314	100.0	1290	6	AAd45219	Aad45219 Human WT
20	2314	100.0	1290	10	ADd84742	Ad84742 Human alp
21	2314	100.0	1290	12	ADJ88268	Adj88268 Human WT
22	2314	100.0	1290	12	ADm48672	Adm48672 Human wil
23	2314	100.0	1290	12	ADN49739	Adn49739 Human alp
24	2314	100.0	1290	13	ADU74415	Adu74415 Human alp
25	2314	100.0	1290	13	ADU66907	Adu66907 Human alp
26	2314	100.0	1290	14	AEA27436	Aea27436 Human alp
27	2314	100.0	1290	14	AEb43187	Aeb43187 Human rGA
28	2314	100.0	1296	6	AAd45222	Aad45222 Human rGA
29	2314	100.0	1296	10	ADd84748	Ad84748 Human alp
30	2314	100.0	1296	12	ADJ88274	Adj88274 Human WT
31	2314	100.0	1296	12	ADm48678	Adm48678 Human wil
32	2314	100.0	1296	13	ADU66913	Adu66913 Human alp
33	2314	100.0	1296	14	AEA27442	Aea27442 Human alp
34	2314	100.0	1306	8	ABZ79777	Abz79777 Human alp
35	2314	100.0	1308	6	AAd45220	Aad45220 Human WT
36	2314	100.0	1308	10	ADd84744	Ad84744 Human alp
37	2314	100.0	1308	12	ADJ88270	Adj88270 Human WT
38	2314	100.0	1308	12	ADm48674	Adm48674 Human wil
39	2314	100.0	1308	13	ADU66909	Adu66909 Human alp
40	2314	100.0	1308	14	AEA27438	Aea27438 Human alp
41	2314	100.0	1319	10	ACF80581	Acf80581 Human alp
42	2314	100.0	1339	10	ADH54464	Adh54464 Human alp
43	2314	100.0	1343	2	AAV31801	Aav31801 Nucleotid
44	2314	100.0	1343	3	AAA75434	Aaa75434 Nucleotid
45	2314	100.0	1343	3	AAA70685	Aaa70685 Human alp

ALIGNMENTS

RESULT 1
AAD45223
ID AAd45223 standard; DNA; 1266 BP.
XX
AC AAd45223;
XX
DT 27-DEC-2002 (first entry)
XX
DE Human rGAL-8 DNA.
XX
KW Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-8; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 1..1266
FT /*tag= a
FT /product= "Human rGAL-8 protein"
XX
PN US2002088024-A1.
XX
PD 04-JUL-2002.
XX
PF 13-NOV-2001; 2001US-00993059.
XX
PR 26-JUL-2000; 2000US-00626127.
XX

PA (GARG/) GARGER S J.
 PA (TURP/) TURPEN T H.
 PA (KUMA/) KUMAGAI M H.
 XX
 PI Garger SJ, Turpen TH, Kumagai MH;
 XX
 DR WPI; 2002-681656/73.
 DR P-PSDB; AAE28210.
 XX
 PT Novel human alpha-galactosidase polypeptide useful for treating lysosomal
 PT storage diseases.
 XX
 PS Claim 1; Page 42-44; 88pp; English.
 XX
 CC The invention relates to human alpha-galactosidase truncated at the
 CC carboxy terminus and the production of enzymatically active recombinant
 CC human and animal lysosomal enzymes. The invention is useful for producing
 CC lysosomal enzymes for treating lysosomal storage diseases, producing
 CC altered or mutated proteins, enzymatically active or otherwise, to serve
 CC as precursors or substrates for further in vivo or in vitro processing to
 CC a specialised industrial form for research or therapeutic uses, to
 CC produce more effective therapeutic enzyme, for producing antibodies
 CC against lysosomal enzymes for medical diagnostic use, and in any
 CC commercial process that involves substrate hydrolysis. The present
 CC sequence is human rGAL-8 DNA
 XX
 SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.15e-237 Length: 1266
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-602-219-12 (1-421) x AAD45223 (1-1266)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGGCTTGGCTTCTCGCTTCTGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 DB 61 CTCGTTCTCGGACATCCCTGGGCTAGAGCACTGGACATGGATTGGCAAGAGCGCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 DB 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCA 180
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 DB 181 GATTCCTGCATCAGTGAGAAGCTTCTCATGGAGATGGCAGAGCTCATGGTCTTCAGAAGGC 240
 QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 DB 241 TGAAGGATGCAAGTTATGAGTACCTCTCATGTGATGACTGTGGATGGCTCCCAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 DB 301 GATTGAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
 QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 DB 361 GCTAAATTATGTTTCACAGCAAAAGGACTGAAGCTAGGGAATTTATGAGATGTTGGAATAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
 DB 421 ACTCGGCAGGCTTCCCTGGGAGTTTGGATGTAATACTAGCATTTGATGCCAGACCTTTGCT 480
 QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 DB 481 GACTGGGGAGTAGATCTGCTGCTAAAAATTTGATGGTTGTTACTGTGACAGATTTGGAAAAATTTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 DB 541 GCAGATGGTTATAGCACATGTCTTGGCCCTGAATAGACTGGCAGAGCAATTTGTGTAC 600
 QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 DB 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 DB 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTGATTCCTGGAAGATATAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGTTGCTGGACAGGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 DB 781 GGTGGATGACCCAGATATGTAGTGATTGGCAACTTTGGCTCAGCTGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTATTTCATGCTTAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 DB 901 CGACACATCAGCCCTCAAGCCCAAGCTCTCTCTCAGGATAGGAGCTAATTTGCCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 DB 961 CAGGACCCCTTGGGCAAGCAAGGTACCACTTAGACAGGGAGAGCAACTTTTGAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValIleMetIleAsnArgGlnGluIleGly 360
 DB 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGGAGGAGATTGGT 1080
 QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 DB 1081 GGACCTCGCTCTTATACCATCGAGTGTCTCCCTGGGTAAGGAGTGGCCTGTATCTCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 DB 1141 GCCTGCTTCATCACACAGCTCTCTCCTGTGAAAAGGAAGCTAGGCTTCTATGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrValLeuLeuGlnLeuGlnAsnThr 420
 DB 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACTGTTTTCCTTCAGCTAGAAAACACA 1260
 QY 421 Met 421
 DB 1261 ATG 1263
 RESULT 2
 ADD84750
 ID ADD84750 standard; DNA; 1266 BP.
 XX
 AC ADD84750;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human alpha-galactosidase rGAL-8 DNA.
 XX
 KW Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
 KW enzyme replacement therapy; lysosomal disease.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1266
 FT /*tag= a
 FT /product= "Human rGAL-8"
 XX

PN US2003106095-A1.
 XX 05-JUN-2003.
 XX 20-MAR-2002; 2002US-00103327.
 XX 26-JUL-2000; 2000US-00626127.
 PR 13-NOV-2001; 2001US-00993059.
 XX (GARG/) GARGER S J.
 PA (TURP/) TURPEN T H.
 XX (KUMA/) KUMAGAI M H.
 XX Garger SJ, Turpen TH, Kumagai MH;
 XX WPI; 2003-801257/75.
 DR P-PSDB; ADD84751.
 XX New polynucleotide for producing active recombinant human and animal
 PT lysosomal enzymes in a plant expression system that can be used in enzyme
 PT replacement therapy.
 XX Claim 1; SEQ ID NO 11; 77pp; English.
 XX The invention relates to human alpha-galactosidase derivatives and the
 CC nucleic acids encoding them. The polypeptides are used in a method for
 CC producing active recombinant human and animal lysosomal enzymes in a
 CC plant expression system. The enzymes can be used in enzyme replacement
 CC therapy for the therapeutic treatment of human and animal lysosomal
 CC diseases. This sequence represents DNA encoding a human alpha-
 CC galactosidase derivative polypeptide of the invention.
 XX
 SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.15e-237 Length: 1266
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-602-219-12 (1-421) x ADD84750 (1-1266)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCTGGCGCTTGCCTTCCTTCTGCCC 60
 Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTCGTTTCTCTGGACATCCCTGGGCTAGAGCACTGGACATGGATTGGCAAGAGCGCT 120
 Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGGCTGGCTGCATCTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180
 Qy 61 AspSerCysIleSerGluLeuLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
 Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAGGATGACGGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
 Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 CATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTCTCTCATGGGATTGGCCAGCTA 360
 Qy 121 AlaAsnTrpValHisSerIlyGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
 Db 361 GCTAATATTGTTCACAGCAAGAGCTGAAGCTAGGAGATTATGCAGATGTTGGAAATAAA 420
 Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyTrpAspIleAspAlaGlnThrPheAla 160

Db 421 ACCTGCGCAGGCTTCCCTGGAGTTTGGATACACGATGTGATGCCACCTTTGCT 480
 Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGGAGTAGATCTGCTAAANATTTGATGGTTTACTCTGACAGATTTGGAAATTTG 540
 Qy 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
 Db 541 GCAGATGGTTATATAGCACATGCTCTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
 Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
 Db 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA 660
 Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTGCATCACTGCGGAAATTTTCTGACATTTGATGATTTCTGGAAAGTATAAAG 720
 Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACGAGGAGAAATTTGTATGTTGTCTGGACAGGG 780
 Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 781 GGTTGGAAATGACCCAGATATATGTTAGTATGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
 Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAATCTAGATGGGCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
 Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGAGCTAATTTGCCATCAAT 960
 Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGGCAAGCAAGGCTTACAGCTTAGACAGGGAGACAACTTTGAAAGTGGG 1020
 Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTAGCTATGATATAACCGCAGGAGATTGGT 1080
 Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCCTGTAAATCCT 1140
 Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyTrpThr 400
 Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAGCTAGGGTTCTATGAATGGACT 1200
 Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 Db 1201 TCAAGTTTAAAGTACATCAATAAATCCCAAGGACTGTTTGTCTTACGTAGAAAACACA 1260
 Qy 421 Met 421
 Db 1261 ATG 1263

RESULT 3
 ADJ88276
 ID ADJ88276 standard; DNA; 1266 BP.
 XX
 AC ADJ88276;
 XX 06-MAY-2004 (first entry)
 XX Human WT rGAL-8 (galactosidase) DNA.
 XX Galactosidase; GAL; gene therapy; lysosomal storage disease;
 KW Fabry's disease; Gaucher's disease; human; gene; ds.
 XX Homo sapiens.

```

XX Key Location/Qualifiers
FH 1. .1266
FT /*tag= a
FT /product= "Human WT rGAL-8 protein"
XX US2004016021-A1.
XX 22-JAN-2004.
XX 23-JUN-2003; 2003US-00602219.
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 15-JUL-1988; 88US-00219279.
XX 17-FEB-1989; 89US-00310881.
XX 05-MAY-1989; 89US-00347637.
XX 08-JUN-1989; 89US-00363138.
XX 22-OCT-1990; 90US-00600244.
XX 16-JAN-1991; 91US-00641617.
XX 26-JUL-1991; 91US-00737899.
XX 01-AUG-1991; 91US-00739143.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX (TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
XX Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-108227/11.
XX P-PSDB; ADJ88277.
XX
XX New lysosomal enzymes, useful in treating human and animal lysosomal
XX storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
XX Claim 1; SEQ ID NO 11; 71pp; English.
XX
XX The invention relates to nucleotide encoding galactosidase (GAL). The
XX invention is useful in gene therapy. The polynucleotides and polypeptides
XX are useful in treating human and animal lysosomal storage diseases, e.g.
XX Fabry's disease and Gaucher's diseases. The present sequence is human GAL
XX DNA.
XX
XX SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: Length: 1266
XX Score: 2314.00 Matches: 421
XX Percent Similarity: 100.00% Conservatives: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 12 Gaps: 0
XX
XX US-10-602-219-12 (1-421) x ADJ88276 (1-1266)
XX
XX 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
XX
XX 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
XX
XX 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAsnGlyLeuAlaArgThrPro 40
XX
XX 61 CTCGTTCTCGGACATCTCCCTGGGCTAGAGCACCTGGACAAATGGATGGCAAGGACGCT 120
XX
XX 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
XX
121 ACCATGGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCAGAGAGCCA 180
61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
181 GATTCTGTCATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
241 TGGAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
301 GATTCAAGAGGACAGACTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
361 GCTAATATGTTTACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAA 420
141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTTACTGTGACAGTTTGGAAATTTG 540
181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
541 GCAGATGGTTATAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
601 TCCGTGAGTGCGCTCTTTATATGTGGCCCTTTCAAGAGCCCAATTATACAGAAATCCGA 660
221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
661 CAGTACTGCAATCACTGGCGAAATTTTGTGCATTTGATGATTCCTCGGAAAGTATAAG 720
241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
721 AGTATCTTTGGACTGGACATCTTTAAACAGAGAGAAATTTGTGATGTTGTCTGGACACGG 780
261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
781 GGTGGAAATGACCCAGATATGTTAGTGTGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTTCATGCTTAATGACCTC 900
301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTCAGGATAAGGACGTAAATTCGCCATCAAT 960
321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
961 CAGGACCCCTTGGCAAGCAGAGGTACAGGTTAGACAGGAGAGACAACTTTGAAAGTGGG 1020
341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTGTATGATAAACCAGCAGAGATTGGT 1080
361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTAGTGGCTGTAATCCT 1140
381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
1141 CCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAAGGAGCTAGGGTTCATGAAATGGACT 1200
401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420

```

Db 1201 TCAGGTTAAGAGTCACATTAATCCCAAGGCACTGTTTGGCTTCAGCTAGAAAACACA 1260

Qy 421 Met 421
|||

Db 1261 ATG 1263

RESULT 4
ADM48680
ID ADM48680 standard; DNA; 1266 BP.
XX
XX
AC ADM48680;
XX
XX 03-JUN-2004 (first entry)
XX
XX Human wild type rGAL-8 DNA.
XX
XX Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW gene; ds.
XX
XX Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT 1..1266
FT /*tag= a
FT /product= "GAL-8 protein"
XX
XX US2004023281-A1.
XX
XX 05-FEB-2004.
XX
XX 23-JUN-2003; 2003US-00602220.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626137.
XX 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX (POGU/) POGUE G P.
XX (ERWI/) ERWIN R L.
XX (GRIL/) GRILL L K.
XX
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
PI P-PSDB; ADM48681.
XX
XX WPI; 2004-142650/14.
XX P-PSDB; ADM48681.
XX
XX New alpha-galactosidase polypeptides, useful in producing recombinant
PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
PT disease.
XX
XX Disclosure; SEQ ID NO 11; 72pp; English.
XX
XX The present invention relates to novel galactosidase (Gal) proteins such
CC as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
CC of the present invention are useful for producing recombinant lysosomal
CC enzymes for enzyme replacement therapy for treating human and animal
CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
CC human wild type rGAL-8 DNA used in the exemplification of the invention.
XX

SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.15e-237 Length: 1266
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-10-602-219-12 (1-421) x ADM48680 (1-1266)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
|||
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCGCTTCGCTTCCTGGCC 60
|||

Qy 21 LeuValSerTrpAspIleProGlyValaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
|||
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGACATGGACATGATGATGGCAAGGACGCT 120
|||

Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
|||
Db 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCGCAGAAAGGCCA 180
|||

Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
|||
Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGAGATGGCAGAGCTCATGCTCTCAGAGGC 240
|||

Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
|||
Db 241 TGGAAGGATGAGGTTATGATCTCTGCAATTCATGACTGTTGGATGGCTCCCAAGA 300
|||

Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyVileArgGlnLeu 120
|||
Db 301 GATTCAGAGGCGACATTCAGCGAGACCTCAGCGCTTCTCTCATGGGATTCGCCAGCTA 360
|||

Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
|||
Db 361 GCTAATTATGTTTCAGCAAAAGAGCTGAAGCTAGGATTTATGCAAGATGTTGGAAATAA 420
|||

Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
|||
Db 421 ACCTGGCGAGGCTTCCCTGGGAGTTTTGGATCTACGATGATGTCGCCAGACCTTTGCT 480
|||

Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
|||
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTTACTCTGCAGCTTTCGAAAATTTG 540
|||

Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
|||
Db 541 GCAGATGGTTATAAGCACATGCTCCTTGGCCCTGAATAGGACTGGCAGAAGCATTTGTGTAC 600
|||

Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
|||
Db 601 TCCGTGAGTGGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCCAATTATACAGAAATCCGA 660
|||

Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
|||
Db 661 CAGTACTGCATCCTGCGGAAATTTTGTGCATTTGATGATTCCTCGGAAAGATATAAG 720
|||

Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
|||
Db 721 AGTATCTTGGACTGGACATCTTTAAACCCAGGAGAGAATTTGTTGTTGTGGACAGGG 780
|||

Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
|||
Db 781 GGTTGGAGTACCAGATATCTTATGTTAGTTGGCACTTTGGCCCTCAGCTGGATCAGCAA 840
|||

Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
|||
Db 841 GTAACCTCATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
|||

Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
|||

Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGCAGTAATTCCTCAAT 960
Qy 321 GlnAspProLeuGlyGlnGlyTyrGlnLeuArgGlnGlyAspAsnPhelValTrp 340
Db 961 CAGSACCCCTCGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluileGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGAGCTATGATATAACCGCAGGAGATTGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTATACCATCGCAGTTGCTTCCTCGGTAAAGAGTGGCCTGTAATCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCCTCCTGTGTAAGAGAAAGCTAGGGTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGGTGAAGAGTCACATAAATCCACAGGCACTGTTTTCTTCAGCTAGAAACACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
RESULT 5
ID ADU66915 standard; DNA; 1266 BP.
AC ADU66915;
XX
DT 10-FEB-2005 (first entry)
XX
DE Human alpha-galactosidase protein encoding DNA #5.
XX
KW Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 1..1266
FT /*tag= a
FT /product= "Human alpha-galactosidase protein"
XX
XX US2004234516-A1.
XX
XX 25-NOV-2004.
XX
XX 21-MAY-2004; 2004US-00851388.
XX
XX 26-JUL-2000; 2000US-00626127.
XX
XX 13-NOV-2001; 2001US-00930359.
XX
XX 20-MAR-2002; 2002US-00103327.
XX
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2004-821274/81.
XX
XX P-PSDB; ADU66916.
XX
XX A pharmaceutical composition comprising a lysosomal enzyme, useful for
XX enzyme replacement therapy for the treatment of lysosomal storage
XX diseases, such as Fabry's disease.
XX
XX Disclosure; SEQ ID NO 11; 88pp; English.
XX
XX The present invention relates to the production of human and animal

CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase enzymes having a post-translational modification provided
CC by the plant expression system. The invention is useful in enzyme
CC replacement therapy for treating lysosomal storage diseases such as
CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase protein encoding DNA.
XX
SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1.15e-237 Length: 1266
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Db: 13 Gaps:
US-10-602-219-12 (1-421) x ADU66915 (1-1266)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACTCTGGCTGGCGCTTGGCTTCCTCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGCGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAAGAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGNAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGACAGGTTATGAGTACCTCTGCATTGATGACTGTGGATTGGCTCCCAAGA 300
Qy 101 AspSerGlyArgGluLeuAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGGACAGACTTCAGGACACCTTCAGCGCTTTCTCATGGGATTGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGAGATTATGCAGATTGTGGAATATA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCTGGGAGTTTTGGATACTACGACATTGATGCCAGACCTTGGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTTTGTCTGTGACAGATTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACATGTCTTGGCCCTGATAGGACTGGCAGAGCAATTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluileArg 220
Db 601 TCCTGTAGTGGCTCTTTATATATGTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTCAATCACTGGCGAAATTTTCTGACATTGATGATTCCTGAAAGATATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

Db 721 AGTATCTGGACTGGACATCTTTAAACAGAGAGAAATTGTGATGTGTGGACCAAGG 780
 Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 781 GGTGGATGACCCAGATATGTAGTATGGCAACTTTGGCCTCAGCTGGAATCAGCA 840
 Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACCTCAGATGGCCCTCTGGCTATCATGCTGCTCTTTATTTCATGCTTAATGACCTC 900
 Qy 301 ArgHisIleSerProGlnAlaIysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTGCCATCAAT 960
 Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGACCCCTTGGCCAGCAAGGTACCACTTAGACAGGAGACACATTTGAAGTGTGG 1020
 Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAAACCGCAGGAGATGGT 1080
 Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCCAGATGTGCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
 Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
 Db 1141 GCCTGCTTCATCACACAGCTCTCTCTCTGTGAAGAGAGCTAGGGTTCATGATGACT 1200
 Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 Db 1201 TCAAGGTAAAGAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTAGAAAAACACA 1260
 Qy 421 Met 421
 Db 1261 ATG 1263
 RESULT 6
 ID AEA27444 standard; DNA; 1266 BP.
 AC AEA27444;
 XX
 DT 11-AUG-2005 (first entry)
 XX
 DE Human alpha-galactosidase DNA, rGAL-8, SEQ ID NO: 11.
 XX
 KW Gauchers disease; metabolic; neurological disease; niemann pick disease;
 KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
 KW antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1. 1266
 FT /ftsg= a
 FT /product= "Human alpha-galactosidase protein"
 XX
 US2005125859-A1.
 FN
 XX
 PD 09-JUN-2005.
 XX
 PP 08-NOV-2004; 2004US-00984389.
 XX
 PR 26-JUL-2000; 2000US-00626127.
 PR 13-NOV-2001; 2001US-00993059.
 PR 20-MAR-2002; 2002US-00103327.
 XX
 PA (LARG-) LARGE SCALE BIOLOGY CORP.

PI Garger SJ, Turpen TH, Kumagai MH;
 DR WPI; 2005-404004/41.
 DR P-PSDB; AEA27445.
 XX
 PT New isolated polypeptides useful for producing lysosomal enzymes in
 PT plants to be utilized in enzyme replacement therapy or for the
 PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
 PT Gaucher's disease.
 XX
 PS Disclosure; SEQ ID NO 11; 88pp; English.
 CC
 CC The present invention relates to the production of human and animal
 CC lysosomal enzymes in plants by a transient plant expression system. The
 CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
 CC galactosidase (Gal) enzymes having a post-translational modification
 CC provided by the plant expression system. The invention is useful in
 CC enzyme replacement therapy for treating lysosomal storage diseases such
 CC as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
 CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
 CC also useful in researches for developing new approaches to medical
 CC treatment of lysosomal storage diseases and in industrial processes
 CC involving enzymatic substrate hydrolysis. The present sequence is the
 CC human alpha-galactosidase DNA.
 XX
 SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,15e-237 Length: 1266
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservations: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 14 Gaps: 0
 US-10-602-219-12 (1-421) x AEA27444 (1-1266)
 Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
 Qy 21 LeuValSerTrpAspIleProGlyValaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120
 Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
 Db 121 ACCATGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCAGGAAGAGCCA 180
 Qy 61 AspSerCysIleSerGlyLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCTGCATCAGTGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 Qy 81 TrpLysAspAlaGlyTrpGluTyrlieuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAAAGGATGCAGGTTATGATGATCTCTGCACTGCTGATGATGCTGCCCAAGA 300
 Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTTCAGAGGCACACTTCAGGAGACCTCAGGCTTCTCTCATGGATTCGCCAGCTA 360
 Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 Db 361 GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAAATAAA 420
 Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACACGATTCATGATGCCACACCTTCTCT 480
 Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGATCTGCTAATAATTTGATGTTTACTGTGACAGTTTGGAAATTTTG 540


```
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGAGGCTTCCCTGGAGTGTGGATACACGACATGTAGTCCAGACCTTGTCT 480
QY 161 AspTrpGlyValAspLeuLeuLeuPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGATTGGAAAAATTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 CGAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGACTGGCAGGAACATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTAGTGGCCCTCTTATATGTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCATCGCGAAATTTGCTGACATGTGATGATTCCTGGAAAGTATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTGGACTGACATCTTTTAACCCAGGAGAAATGTTGATGTTGCTGGACCAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGAAATGACCCAGATGATGTAGTATGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaIleProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAGGACCTAATGGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTACACAGCTTAGACAGGGAGACAACTTTGAAAGTGTG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACACCTCTCTCAGGCTAGCCCTGGGCTGTAGCTATGATAAACCCGACGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaLaserLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTGCTCCCTGGGTAAAGGAGTGGCCCTGTAATCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGGAAGCTAGGCTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db 1201 TCAAGTTTAAAGATGCATATAATCCACAGGCACTGTTTGTCTTTCAGCTAGAAAACACA 1260
QY 421 Met 421
Db 1261 ATG 1263
RESULT 8
ADD84746
ID ADD84746 standard; DNA; 1278 BP.
XX
AC ADD84746;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human alpha-galactosidase rGAL-4 DNA.
XX
KW Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease.
XX
```

```
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 1..1278
FT /*tag= a
FT /product= "Human rGAL-4"
XX
PN US2003106095-A1.
XX
PD 05-JUN-2003.
XX
PF 20-MAR-2002; 2002US-00103327.
XX
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX
Garger SJ, Turpen TH, Kumagai MH;
XX
WPI; 2003-801257/75.
P-FSDB; ADD84747.
XX
New polynucleotide for producing active recombinant human and animal
PT lysosomal enzymes in a plant expression system that can be used in enzyme
PT replacement therapy.
XX
PS Claim 1; SEQ ID NO 7; 77pp; English.
XX
CC The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for
CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents DNA encoding a human alpha-
CC galactosidase derivative polypeptide of the invention.
SQ Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.17e-237 Length: 1278
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-602-219-12 (1-421) x ADD84746 (1-1278)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTCTCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyValaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGCTTCTCTGGACATCCCTGGGCTAGAGCACTGGACATGATGATGGCAAGGAGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGATGACAGTTATGATGACTCTGCAATTGATGATGATGATGATGATGATGATG 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
XX
```

```
Db      301 GATTGAGAGGAGAGCTTCAGGACAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy      121 AlaAsnTyrValHisSerIysGlyLeuIysLeuGlyIleTyrAlaAspValGlyAsnIys 140
      |||
Db      361 GCTAATATGTTACACAGCAAGAGCTGAAGCTAGGGATTTATGCAAGATGTTGGAATAAAA 420
Qy      141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
      |||
Db      421 ACCTGGCAGGGCTTCCTGGGAGTTTGGATGACTAGCAATGATGCCAGACCTTTGCT 480
Qy      161 AspTrpGlyValAspLeuLeuIysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
      |||
Db      481 GACTGGGAGTAGACTGCTAAAATTTGATGTTGTTACTGTGACAGATTTGGAAAATTTG 540
Qy      181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
      |||
Db      541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAGCAATGTGTAC 600
Qy      201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
      |||
Db      601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTCAAAGCCCAATTTATACAGAAATCCGA 660
Qy      221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpIysSerIleIys 240
      |||
Db      661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAGTATAAG 720
Qy      241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
      |||
Db      721 AGTATCTTGGACTGGACATCTTTTAACACGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
Qy      261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
      |||
Db      781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCTTACGCTGGGAATCAGCAA 840
Qy      281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
      |||
Db      841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTTATTCATGCTTAATGACCTC 900
Qy      301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
      |||
Db      901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
Qy      321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
      |||
Db      961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTTAGACAGGAGACAACTTTCAAGTGTGG 1020
Qy      341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
      |||
Db      1021 GAACGACCTCTCTCAGGCTTAGCGTGGGCTGTAGCTATGATAAAACCGGCGAGAGATTGGT 1080
Qy      361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
      |||
Db      1081 GGACCTCGCTCTTATACCATCGAGTGTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
Qy      381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
      |||
Db      1141 GCTGTGTTTCATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
Qy      401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
      |||
Db      1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260
Qy      421 Met 421
      |||
Db      1261 ATG 1263

RESULT 9
ID      ADJ88272 standard; DNA; 1278 BP.
XX
AC      ADJ88272;
XX
DT      06-MAY-2004 (first entry)
```

```
XX      Human rGAL- 4 (galactosidase) DNA.
DE
XX      Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW      Fabry's disease; Gaucher's disease; human; gene; ds.
XX
OS      Homo sapiens.
XX
FH      Key      Location/Qualifiers
      CDS      1..1278
FT      /*tag= a
FT      /product= "Human WT rGAL-4 protein"
FT      /partial
FT      /note= "No start codon"
XX
PN      US2004016021-Al.
XX
PD      22-JAN-2004.
XX
XX      23-JUN-2003; 2003US-00602219.
XX
PR      26-FEB-1988; 88US-00160766.
PR      26-FEB-1988; 88US-00160771.
PR      15-JUL-1988; 88US-00219279.
PR      17-FEB-1989; 89US-00310881.
PR      05-MAY-1989; 89US-00347637.
PR      08-JUN-1989; 89US-00363138.
PR      22-OCT-1990; 90US-00600244.
PR      16-JAN-1991; 91US-00641617.
PR      26-JUL-1991; 91US-00737899.
PR      01-AUG-1991; 91US-00739143.
PR      31-JUL-1992; 92US-00923692.
PR      30-DEC-1992; 92US-00997733.
PR      29-DEC-1993; 93US-00176414.
PR      19-JAN-1994; 94US-00184237.
PR      14-OCT-1994; 94US-00324003.
PR      21-MAY-1999; 99US-00316572.
PR      26-JUL-2000; 2000US-00626127.
PR      13-NOV-2001; 2001US-00993059.
XX
PA      (TURP/) TURPEN T H.
PA      (POGU/) POGUE G P.
PA      (ERWI/) ERWIN R L.
PA      (GRIL/) GRILL L K.
XX
PI      Turpen TH, Pogue GP, Erwin RL, Grill LK;
DR      WPI; 2004-108227/11.
DR      P-PSDB; ADJ88273.
XX
PT      New lysosomal enzymes, useful in treating human and animal lysosomal
      storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
PS      Claim 1; SEQ ID NO 7; 71pp; English.
XX
CC      The invention relates to nucleotide encoding galactosidase (GAL). The
      invention is useful in gene therapy. The polynucleotides and polypeptides
      are useful in treating human and animal lysosomal storage diseases, e.g.
      CC      Fabry's disease and Gaucher's diseases. The present sequence is human GAL
      CC      DNA.
XX
SQ      Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:      1.17e-237      Length:      1278
Score:          2314.00      Matches:      421
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              12      Gaps:      0
```

US-10-602-219-12 (1-421) x ADJ88272 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTGGTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCTGCATCAGTCAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysValLeuAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAAAGGATGCAAGGTTATGAGTACTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTGAGAGGAGAGCTTCAGGAGAGCTTCAGGCTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaSerTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 361 GCTAATTATGTTACAGCAAGAGCTGAAGCTAGGAGTATATGACAGATGTTGGAAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGCGGAGGCTCCCTGGGAGTTTGGATCTACGACATTTGATGCCGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGATCTCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTG 540
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
DB 541 GCAGATGTTTAAAGCACATGCTCTGGCCCTGAAATAGGACTGGCAGAAGCATTTGTGAC 600
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
DB 601 TCCTGTGAGTGCCCTCTTTATATGTTGGCCCTTTCAAAAGGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCATCATCTGGCGAAATTTTGTGACATTTGATGATCTCTGGAAAGTATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACGAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGAAATGCCAGATATGTTAGTGTATGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACCTAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGAGCTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGGCAAGCAGGGTACCAGCTTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCAGGCTTACGCTTGGGCTGTAGCTATGATAAACCGGACAGAGATTGCT 1080
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380

DB 1081 GGACCTCGCTCTTATACATCGCAGTTGCTTCCTCGGTAAAGGAGTGGCTGTAACTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
DB 1141 GCCTGCTTCATCACACAGCTCTCTCTGTGAAAGGAAGCTAGGTTCTATGATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
DB 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTACGTAGAAAAACA 1260
QY 421 Met 421
DB 1261 ATG 1263
RESULT 10
ADM48676
ID ADM48676 standard; DNA; 1278 BP.
XX
AC ADM48676;
XX
XX 03-JUN-2004 (first entry)
XX Human wild type rGAL-4 DNA.
XX Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW gene; ds.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 4..1278
FT /*tag= a
FT /product= "GAL-4 protein"
FT /partial
FT /note= "No start codon"
XX
PN US2004023281-A1.
PD
PD 05-FEB-2004.
XX
PF 23-JUN-2003; 2003US-00602220.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
PI Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-142650/14.
DR P-PSDB; ADM48677.
XX
XX New alpha-galactosidase polypeptides, useful in producing recombinant
PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
PT disease.

XX Disclosure; SEQ ID NO 7; 72pp; English.

XX The present invention relates to novel galactosidase (Gal) proteins such

CC as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions

CC of the present invention are useful for producing recombinant lysosomal

CC enzymes for enzyme replacement therapy for treating human and animal

CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick

CC disease, Fabry disease and Tay-Sachs disease. The present sequence is

CC human wild type rGAL-4 DNA used in the exemplification of the invention.

XX

SQ Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,17e-237 Length: 1278

Score: 2314.00 Matches: 421

Percent Similarity: 100.00% Conservatives: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 12 Gaps: 0

US-10-602-219-12 (1-421) x ADM48676 (1-1278)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20

Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60

Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40

Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCCTGGCAATGGATTGGCAAGGACGCT 120

Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60

Db 121 ACATGGGCTGGCTGCATCTGGAGGCGCTTCATGTGCACCTTGACTGCCAGGAAGCCA 180

Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80

Db 181 GATTCTCGCATCAGTGAGAAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

Qy 81 TrpLysAspAlaGlyTyrgluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100

Db 241 TGGAAAGGATGCAAGGTATAGTACCTCTGCATGTATGACTGTGGATGGCTCCCAAGA 300

Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120

Db 301 GATTTCAGAGGCGAGACTTCAGGAGACCCCTCAGCGCTTTCCTATGGGATTCGCCAGCTA 360

Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140

Db 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGGATTTATGCAGATGTTGGAAATAA 420

Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160

Db 421 ACCTGCGCAGGCTTCCTCGGAGTTTGTGATATACGACATTTGATGCTCCAGACCTTTGCT 480

Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180

Db 481 GACTGGGAGTAGAUCTGCTAAAATTGTATGTTGTACTGTGACAGCTTTGGAAATTTG 540

Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200

Db 541 CGAGATGGTTAAGCACATGCTCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600

Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220

Db 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660

Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240

Db 661 CAGTACTGCANTACCTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATAAG 720

Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnArgIleValAspValAlaGlyProGly 260

Db 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;

Db 721 AGTATCTGGAGCTGGACATCTTTTAAACCAGGAGAGAAATGTTGATGTTGCTGGACAGG 780

Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280

Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGAACCTTTGGCTCAGCTGGGAATCAGCAA 840

Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300

Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCTATGCTAATGACCTC 900

Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaLeuAsn 320

Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCCTCATCAAT 960

Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340

Db 961 CAGGACCCCTTGGGCAAGCAGGGTACCAGCTTTAGACAGGAGACAACTTTGAAAGTGTGG 1020

Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360

Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCCGCGCAGGAGATTGCT 1080

Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380

Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT 1140

Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400

Db 1141 GCCTGCTCATCATCAGAGCTCTCTCTGTGAAAGAGCTAGGCTTTCTATGATGGACT 1200

Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420

Db 1201 TCAAGTTTAAAGAGTCACATAAATCCACAGGCACTGTTTCTTCTTTCAGTAGAAAAACACA 1260

Qy 421 Met 421

Db 1261 ATG 1263

RESULT 11

ADU66911

ID ADU66911 standard; DNA; 1278 BP.

XX

AC ADU66911;

DT 10-FEB-2005 (first entry)

XX

DE Human alpha-galactosidase protein encoding DNA #3.

XX

KW Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;

KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;

KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;

KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human; gene; ds.

XX

OS Homo sapiens.

XX

Key Location/Qualifiers

CDS 1..1278

FT /tag= a

FT /product= "Human alpha-galactosidase protein"

FT /transl_except= (pos:1..6, aa:Gln)

XX

PN US2004234516-A1.

XX

PD 25-NOV-2004.

XX

PF 21-MAY-2004; 2004US-00851388.

XX

PR 26-JUL-2000; 2000US-00626127.

PR 13-NOV-2001; 2001US-00993059.

PR 20-MAR-2002; 2002US-00103327.

XX

PA (LARG-) LARGE SCALE BIOLOGY CORP.

XX

PI Garger SJ, Turpen TH, Kumagai MH;
 DR WPI; 2004-821274/81.
 DR P-PSDB; ADU66912.
 XX
 XX A pharmaceutical composition comprising a lysosomal enzyme, useful for
 PT enzyme replacement therapy for the treatment of lysosomal storage
 PT diseases, such as Fabry's disease.
 XX
 XX Disclosure; SEQ ID NO 7; 89pp; English.
 XX
 CC The present invention relates to the production of human and animal
 CC lysosomal enzymes in plants by a transient plant expression system. The
 CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
 CC galactosidase enzymes having a post-translational modification provided
 CC by the plant expression system. The invention is useful in enzyme
 CC replacement therapy for treating lysosomal storage diseases such as
 CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
 CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
 CC also useful in researches for developing new approaches to medical
 CC treatment of lysosomal storage diseases and in industrial processes
 CC involving enzymatic substrate hydrolysis. The present sequence is the
 CC human alpha-galactosidase protein encoding DNA.
 XX
 SQ Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,178-237 Length: 1278
 Scores: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 13 Gaps: 0

US-10-602-219-12 (1-421) x ADU66911 (1-1278)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGAACCCAGAACATACATCTGGGCTGGGCTTGGCTTCCTGGGCC 60
 Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTCGTTCTCTGGGACATCCCTGGGCTAGAGCACTGCAATGGATTGGCAAGGAGCGCT 120
 Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTCAACCTTGACTGCAGGAGAGCCA 180
 Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTGTGGATGGCTCCCAAGA 300
 Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTGAGAGGAGCACTTCAGGACAGCCCTCAGCGCTTTCCTCATGGGATGGCCAGCTA 360
 Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
 Db 361 GCTAATTATGTTACAGCAAGAGGACTGAAGCTAGGATTTATGCAGATGTTGGAAATAAA 420
 Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTCGCGCAGGCTTCCTCGGAGTGTGGATACAGCACTTGTGATGGCCAGACCTTTGCT 480
 Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGATCTGCAAAATTTGATGTTGTACTGTGACAGATTGGAAATTTG 540
 Qy 181 AlaAspGlyTyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200

Db 541 GCAGATGGTTATAAGCACTGCTCTGGCCCTGTAATAGGACTGGCAGAGCATTTGTGTAC 600
 Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
 Db 601 TCCGTGTAGTGGCTCTTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
 Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTGCAATCACTGGCGAAATTTTCTGCATTTGATGATTCCTGCGAAAGTATAAAG 720
 Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATTTGTGGACAGGG 780
 Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 781 GGTGGATGATGACCCAGATATGTTAGTGTATGGCACTTTGGCCCTCAGCTGGAAATCAGNA 840
 Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCTATGCTTAATGACCTC 900
 Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAAATTGCCCATCAAT 960
 Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGAGACAACTTTGAACTGTGG 1020
 Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGCAGAGATTGT 1080
 Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
 Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
 Db 1141 CCTGCTTCATCACACAGCTCTCTCTGTGAAAGGAGCTAGGGTCTTATGAATGGACT 1200
 Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 Db 1201 TCAAGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTTTCAGCTAGAAAAACACA 1260
 Qy 421 Met 421
 Db 1261 ATG 1263

RESULT 12
 ID AEA27440 standard; DNA; 1278 BP.
 XX AEA27440;
 AC AEA27440;
 DT 11-AUG-2005 (first entry)
 XX Human alpha-galactosidase DNA, rGAL-4, SEQ ID NO: 7.
 DE Human alpha-galactosidase DNA, rGAL-4, SEQ ID NO: 7.
 XX Gauchers disease; metabolic; neurological disease; niemann pick disease;
 KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
 KW antileptic; cns-gen.; lysosome storage disease; alpha-galactosidase;
 KW gene; ds.
 XX Homo sapiens.
 XX Location/Qualifiers
 FH 1..1278
 FT /*tag= a
 FT /product= "Human alpha-galactosidase protein"
 FT 4..1278
 CDS

```
FT      /*tag= b
FT      /product= "Human alpha-galactosidase protein"
FT      /partial
FT      /note= "No start codon"
```

US2005125859-A1.

09-JUN-2005

08-NOV-2004: 2004US-00984389.

26-JUL-2000: 2000US-00626127.

15 NOV 2001, 2001US-00055035:
20-MAR-2002: 2002US-00103327:

(LARG-) LARGE SCALE BIOLOGY C

Gardner SJ, Turpen TH, Kumaqai M

WPI: 2005-404004/41.

[illegible]

plants to be utilized in en-

Gaucher's disease.

Disclosure; SEQ ID NO 7; 88pp; English.

The present invention relates to the production of human and animal

invention relates to glucocerebrosidase (GCB, GCR) and alpha-

gastrointestinal (GI) enzymes having a poor transmembrane localization provided by the plant expression system. The invention is useful in

as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs

also useful in researches for developing new approaches to medical

involving enzymatic substrate hydrolysis. The present sequence is the

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

THE

No.:	Length:	1278
1.17e-237		

Percent Similarity: 100.00%
Conservative: 0

Match:	100.00%	Indels:	0
--------	---------	---------	---

100

610

THE UNIVERSITY OF CHICAGO

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

AC AAD45224;
 XX 27-DEC-2002 (first entry)
 XX DE Human rGAL-8R DNA.
 XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
 KW therapeutic; rGAL-8R; gene; ds.
 XX OS Homo sapiens.
 XX Key Location/Qualifiers
 XX CDS 1..1284
 FT /*tag= a
 FT /product= "Human rGAL-8R protein"
 XX US2002088024-A1.
 XX 04-JUL-2002.
 XX 13-NOV-2001; 2001US-00993059.
 XX 26-JUL-2000; 2000US-00626127.
 XX (GARG/) GARGER S J.
 XX (TURP/) TURPEN T H.
 XX (KUMA/) KUMAGAI M H.
 XX Garger SJ, Turpen TH, Kumagai MH;
 XX WPI; 2002-681656/73.
 DR P-PSDB; AAE28211.
 XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal
 PT storage diseases.
 XX Claim 1; Page 44-46; 88pp; English.
 XX The invention relates to human alpha-galactosidase truncated at the
 CC carboxy terminus and the production of enzymatically active recombinant
 CC human and animal lysosomal enzymes. The invention is useful for producing
 CC lysosomal enzymes for treating lysosomal storage diseases, producing
 CC altered or mutated proteins, enzymatically active or otherwise, to serve
 CC as precursors or substrates for further in vivo or in vitro processing to
 CC a specialised industrial form for research or therapeutic uses, to
 CC produce more effective therapeutic enzyme, for producing antibodies
 CC against lysosomal enzymes for medical diagnostic use, and in any
 CC commercial process that involves substrate hydrolysis. The present
 CC sequence is human rGAL-8R DNA
 XX Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 1.17e-237 Length: 1284
 Score: 2314.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-10-602-219-12 (1-421) x AAD45224 (1-1284)
 QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 DB 1 ATGCAGCTGAGGAACCCAGAACTACTCTGGGCTGGCGCTTGGCTTCTCGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 DB 61 CTGGTTCTCTGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGAGCGCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 DB 121 ACCATGGGCTGGCTGCACTGGAGGCGCTTCATGTGCAACCTTGTGACGTCAGGAAGAGCCA 180
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 DB 181 GATTCTCGCATCAGTGAAGACTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 DB 241 TGGAAAGGATGAGGTATGATGACTCTCGCATTTGATGCTGTGGATGGCTCCCAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 DB 301 GATTCAAGAGGAGACTTCAGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
 QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 DB 361 GCTAATTATGTTTCACAGCAAGAGCTGAAGCTAGGGATTTATGCAGATGTTTGGAAATAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 DB 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACACGACATTTGATGCCAGACCTTTTGTCT 480
 QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
 QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 DB 541 GCAGATGGTTATAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAAGCATTTGTGTAC 600
 QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
 DB 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCA 660
 QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 DB 661 CAGTACTGCATCACTGCGGAATTTTGTGCATTTGATGATTCCTCGAANAAGTATTAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 DB 781 GGTGGGAATGACCCAGATATGTTAGTATGGCACTTTGGCCCTCAGCTGGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
 DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGGCCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 DB 961 CAGGACCCCTTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGTGGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
 QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 DB 1081 GGACCTCGCTTATACCATCGAGTGTCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 DB 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAGCTAGGGTTCCTATGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
 DB 1201 TCAAGGTTAAGAAGTACATAAATCCCAACAGGCACTGTTTTGCTTCAGCTAGAAAACACA 1260

QY	421 Met 421	Db	CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT	120
DB	1261 ATG 1263	QY	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
RESULT 14		DB	ACCATGGGCTGGCTGGCACTGGGAGGCTTCATGTGCAACCTTGACTGGCAGGAAGGCCA	180
ID	ADD84752 standard; DNA; 1284 BP.	QY	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
XX	ADD84752;	DB	GATTCTCATCATCAGTGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAGGC	240
DT	29-JAN-2004 (first entry)	QY	TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg	100
XX	Human alpha-galactosidase rGAL-8R DNA.	DB	TGGAAGGATGCAGGTTATGAGTACCTCTGCTGATTGATGACTGTGTGATGGCTGCCCAAGA	300
DE	Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;	QY	AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
KW	enzyme replacement therapy; lysosomal disease.	DB	GATTTCAGAAGGCAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA	360
OS	Homo sapiens.	QY	AlaIleTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys	140
FX	Key	DB	GCTAATTATGTTACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAATAA	420
FT	CDS	QY	ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla	160
FT	1..1284	DB	ACCTGGCAGGCTTCCTGGGAGTTTGGATACTAGACATTTGATGCCAGACCTTTGCT	480
FT	/*tag= a	QY	AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu	180
FT	/product= "Human rGAL-8R"	DB	GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGCAGCTTTGGAAATTTG	540
XX	US2003106095-A1.	QY	AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr	200
XX	05-JUN-2003.	DB	GCAGATGGTTATAAGCACATGTCCTGGCCCTGAATAGACTGGCAGAGCAATGTTGTAC	600
XX	20-MAR-2002; 2002US-00103327.	QY	SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg	220
XX	26-JUL-2000; 2000US-00626127.	DB	TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGA	660
XX	PR 13-NOV-2001; 2001US-00993059.	QY	GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys	240
XX	(GARG/) GARGER S J.	DB	CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTGATGATTCCTGGAAAAGTATAAG	720
PA	(TURP/) TURPEN T H.	QY	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly	260
PA	(KUNA/) KUMAGAI M H.	DB	AGTATCTTGGACTGGACATCTTTTAACGAGAGAGATTTGTGATGTTGCTGGACAGGG	780
XX	Garger SJ, Turpen TH, Kumagai MH;	QY	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln	280
XX	WPI; 2003-801257/75.	DB	GGTTGGAATGACCCAGATATGTTAGTATTGGCAACTTTGGCCTCAGCTGGAAATCAGCAA	840
DR	P-PSDB; ADD84753.	QY	ValThrGlnMetAlaLeuTrpAlaIleMetAlaIaProLeuPheMetSerAsnAspLeu	300
XX	New polynucleotide for producing active recombinant human and animal	DB	GTAACCTCAGATGGGCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTTCTAATGACCTC	900
PT	lysosomal enzymes in a plant expression system that can be used in enzyme	QY	ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn	320
PT	replacement therapy.	DB	CGACATCAGCCCTCAAGCCCAAGCTCTCCTTCAGGATAGGAGCTAATGCCATCAAT	960
XX	Claim 1; SEQ ID NO 13; 77pp; English.	QY	GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp	340
CC	The invention relates to human alpha-galactosidase derivatives and the	DB	CAGGACCCCTGGGCAAGCAAGGTTACCCAGCTTAGACAGGGAGACAACCTTCAAGTGTGG	1020
CC	nucleic acids encoding them. The polypeptides are used in a method for	QY	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly	360
CC	producing active recombinant human and animal lysosomal enzymes in a	DB	GAAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCGCAGAGATTGGT	1080
CC	plant expression system. The enzymes can be used in enzyme replacement	QY	GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro	380
CC	therapy for the therapeutic treatment of human and animal lysosomal	DB	GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAACTCT	1140
CC	diseases. This sequence represents DNA encoding a human alpha-	QY	AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr	400
CC	galactosidase derivative polypeptide of the invention.			
XX	Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;			
SQ				
Alignment Scores:				
Pred. No.:	1.17e-237	Length:	1284	
Score:	2314.00	Matches:	421	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	100.00%	Indels:	0	
DB:	10	Gaps:	0	
US-10-602-219-12 (1-421) x ADD84752 (1-1284)				
QY	1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20		
DB	1 ATGCAGCTGAGGAACCCGAACATACATCTGGGCTGGCGCTTGGCGCTTCCTGGCC	60		
QY	21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40		

Db 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGTAAAGAGAGCTAGGTTCTTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAenProThrGlyThrValLeuLeuGlnLeuGluAenThr 420
Db 1201 TCAAGTTTAAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGTAGAACAACA 1260
Qy 421 Met 421
Db 1261 ATG 1263
RESULT 15
ID ADJ88278 standard; DNA; 1284 BP.
XX AC ADJ88278;
XX 06-MAY-2004 (first entry)
XX Human WT rGAL-8 R (galactosidase) DNA.
XX Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease; human; gene; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
FH 1..1284
FT CDS /*tag= a
FT /product= "Human WT rGAL-8R protein"
XX US2004016021-Al.
XX 22-JAN-2004.
XX 23-JUN-2003; 2003US-00602219.
XX 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX (TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX Turpen TH, Pogue GP, Erwin RL, Grill LK;
PI P-PSDB; ADJ88279.
XX WPI; 2004-108227/11.
XX P-PSDB; ADJ88279.
XX New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX Claim 1; SEQ ID NO 13; 71pp; English.
XX The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.

CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
CC DNA.
XX
SQ Sequence 1284 BP; 330 A; 293 C; 335 G; 326 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1.17e-237 Length: 1284
Score: 2314.00 Matches: 421
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0
US-10-602-219-12 (1-421) x ADJ88278 (1-1284)
Qy 1 MetGlnLeuArgAenProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGAACCCAGAACTACATCTGGCTGCGCGCTTGCCTTCCTCTCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyValaArgAlaLeuAspAenGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAenLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTAGAGAGCTCTTCATGAGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGGATGAGGTTATGAGTACCTCTGCTGATGACTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGAGCAGACTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAenTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAenLys 140
Db 361 GCTAATTAATGTTTCAGCAAGAGGACTGAAGCTAGGAGTTATGCAAGATGTTGGAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGTGATACAGACATTTGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAenLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTTGTATGTTTGTACTGTGACAGATTTGGAAAATTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAenArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTATAAGCACATGTCTTGGCCCTGAATAGACTGGCAGAGCAATGTTGTATC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAenTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCCTCTTTATATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAenHisTrpArgAenPheAlaAspIleAspAspSerTrpIlySerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTTGATGATTCCTGGAAGAATATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAenGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTGGACTGGACATCTTTTAACAGAGAGAAATGTTGATGTTGTGGACAGGG 780
Qy 261 GlyTrpAenAspProAspMetLeuValIleGlyAenPheGlyLeuSerTrpAenGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTATTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAenAspLeu 300

```
Db      841  GTAAC TCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTATCATGTCTAATGACCTC 900
Qy
301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      901  CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATTGCCATCAAT 960
Qy      321  GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db      961  CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGGAGACAACTTTGAGTGTGG 1020
Qy      341  GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      1021  GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGCAGGAGATTGCT 1080
Qy      361  GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1081  GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
Qy      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db      1141  GCCTGCTTCATCACAGCTCTCTCCTGTGAAAAGGAGCTAGGGTTCTATGAATGGACT 1200
Qy      401  SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeuGluAsnThr 420
Db      1201  TCAAGGTTAAGAAGTCACATAAATCCCA CAGGCACCTGTTTTGCTTCAGCTAGAAAACACA 1260
Qy      421  Met 421
Db      1261  ATG 1263
```

Search completed: December 26, 2005, 11:08:34
Job time : 683 secs